FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

A Comparison on Statistical Methods and Long Short Term Memory Network Forecasting the Demand of Fresh Fish Products

Rúben Alexandre da Fonseca Marques

DISSERTATION

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Professor Gonçalo Figueira

Co-Supervisor: Professora Vera Miguéis

February 24, 2020

A Comparison on Statistical Methods and Long Short Term Memory Network Forecasting the Demand of Fresh Fish Products

Rúben Alexandre da Fonseca Marques

Mestrado Integrado em Engenharia Informática e Computação

February 24, 2020

Abstract

Fresh fish has a fast deterioration process and the orders must be schedule daily. An accurate estimation of the daily demand in the supply chain help management prevents the fish spoiling and eventually raises profitability.

Historically the industry has relied on statistical models to forecast the demand in a time series approach, but with the raise of computational power, machine learning starts to become a viable option to model time series as well. This thesis aims to be a contribution to validate the accuracy of neural networks modelling time series demand.

The results of the study shows that a Long Short Term Memory network can be modeled to forecast the demand with the same accuracy of the traditional models, but requires much effort in the preprocessing of the data. But that effort is rewarded with an accurate forecast of the demand. However, the statistical methods has a straightforward implementation with similar accuracy and should not be discarded as a viable option.

Acknowledgements

Esta dissertação marca o ponto final do meu percurso no Mestrado Integrado em Engenharia Informática e Computação na FEUP. Deste modo, encerro um capítulo da minha vida repleto de histórias, dificuldades ultrapassadas e de momentos inesquecíveis. Da FEUP levarei as memórias de um caminho feito de aprendizagem que me ajudou imenso no meu crescimento pessoal. Serei eternamente grato à FEUP e a todas as pessoas que dela fazem parte.

Um agradecimento especial e merecido aos meus orientadores. O professor Gonçalo Figueira e a professora Vera Miguéis, que mostraram uma disponibilidade total para ajudar e para discutir o melhor caminho para a resolução dos vários desafios que foram surgindo durante a dissertação.

Aos amigos que comigo percorreram este percurso e que contribuiram para que eu chegasse a este momento, a todos voçes um obrigado.

Agradeço aos meus pais, José e Arlete, por todo o apoio que me deram durante este percurso, que sem voçes não seria possível.

Por fim, agradeço à Liliana. Por todo o amor nos bons momentos e apoio quando eu mais precisava.

Rúben Marques

Contents

List of Figures

List of Tables

Abbreviations

ANN	Artificial Neural Network
POS	Points of sale
MLP	MultilayerPerceptron
SVM	Support Vector Machine
RBFN	Radial Basis Function Network
SARIMA	Seasonal Autoregressive Integrated Moving Average with External Variables
HW	HoLt Winters
LSTM	Long Short Term Memory
RMSE	Root mean Square Error
NRMSE	Normalized Root Mean Square Error
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
TBATS	Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components
SKU	Stock keeping Unit
RNN	Recurrent Neural Networks
ACF	Auto correlation function

Chapter 1

Introduction

1.1 General Context

Retail industry is involved in distributing goods from the producer to the consumers, using physical locations like stores or more recently, providing the service on online channels[MKF18]. Retailers can be small shops that are part of a community to international companies that have stores all around the world. Such diversity gives the consumer a lot to choose and is essential for companies to adopt strategies to retain costumers and attract new ones.

The competition in the industry is fierce and any competitive advantage over rivals company, is crucial to overcome such a challenging market. The use of data to improve business operations, become a major issue for big retailers[SD17]. One of those cases is to sales forecast. A reliable estimation of sales, should be key to understand the demand by the consumer and provides distribution, profitability and reduce holding inventory and waste.

Portugal has a long tradition of fresh fish food, from consumption to fishing. Fishing has been a way of life in Portugal for generations, providing employment and natural resource that is part of our diet[CLC+13]. For both reasons, is essential find strategies that help our resources to be sustainable. Forecasting fresh food sales will improve sustainability because will provide additional knowledge in consumption patters and reduce waste.

1.2 Problem Description

This dissertation has the partnership of a retail corporation that operates a chain of hypermarket and provides groceries to satisfy demand through the supply chain. The retailer supply multiple products, but we are concerned in forecasting the demand of fresh fish products. Since fresh fish has a fast deterioration it can't be stored for a long period of time like other products and the schedule of orders must be done daily, so the freshness is always available to satisfy the customer.

The fish are supplied to the stores trough two different channels of the supply chain, the aquaculture and wild caught industry. The previous scheduling of orders, must be done according to the forecast of the demand, to ease the responsibility of manage hundreds of products in hundreds of stores. Without the proper tools, this can become a excruciating task.

The retail partner collects the data from the transactions in the stores, that becomes potential sources of information. By extracting useful information from that data, the modeling of an accurate forecasting model will prevent spoiled fish and stockouts, resulting in less waste and an increase of profitability.

1.3 Objectives and methodology

The main objective of this dissertation is to develop a model to forecast the demand of fresh fish product in a two day ahead horizon. To accomplish this objective a Long Short Term Memory(LSTM) will be develop with multiple features as input, such as price, meteorologic or holidays. The result of the LSTM will need validation, so statistical models will be also implemented as verification of the accuracy of the machine learning model.

The comparison of the implemented models results, will be a contribution to validate or disprove the modeling of time series using neural networks. The improvement on computational power in recent years, made machine learning a viable option to model time series, but more results are required to consolidate the evidence.

The methodology can be described as follows:

Exploration of the data: The preprocessing of the transactional dataset provided the the retail partner, to identify the most valuable information worth to extract. Along with the transactional data, research evidence in the literature of predictive features that already had results.

Development of the models: The implementation of multiple forecasting models will benefit the end result.

Validate the results: A comparison of the results, to evaluate the accuracy of the develop models.

1.4 Thesis Outline

This thesis includes 2 chapters, which are summarized below.

Chapter 2 intends to provide theoretical background on forecasting, a general description on retail forecasting and specifically a extend review on sales and product demand forecasting.

Chapter 3 describes the forecasting methods to be implemented during the course of this dissertation.

Chapter 4 describes the problem description and presents the process of the exploration of the transactional data.

In chapter 5, the forecast of the fresh fish demand is performed. The implementation and results of the statistical methods is displayed. In the second part, the LSTM process and feature selection is detailed, ending with the comparison of results with the statistical models. The previous models, are also applied to a smaller dataset to validate results with a comparison with the retail's partner model.

Chapter 6 presents the summary and conclusions of the research developed in this thesis. It also presents some directions for future research.

Chapter 2

Literature Review on Forecasting

Forecasting is the process of predicting the future, using past information. Intuitively, we do it everyday taking decisions using our life experience, expertise or emotions. This type of forecasting is qualitative and it is not objective. Forecasting can also be quantitative, where the possibility of a future event occurring is estimated using a modeled approach. The base idea, is the belief that historical data is cyclic and patterns or trends are repeated over time.

The traditional approach of quantitative forecasting is by statistical models, but with the technological improvement of computational power, machine learning turned out to be an option. In machine learning, a computer system learns information about a topic, by being fed with observations about it. Despite machine learning being an option to forecasting, its performance must be validated. Studies comparing models of machine learning and statistical methods[AQS01][MSA18][CZ03][ABGM15], had mixed results, with some giving the advantage to machine learning, and vice versa. Regarding the results, both have great potential for forecasting applications [MSA18], with each one having a better performance depending on a variety of characteristics. Therefore, it is important that we classify the different forecasting settings and their challenges. This is explored in the following sections.

Forecast is useful to provide us support to decisions and is used and studied in many areas, such as weather[WSH99][Gne05], stock market[HNW05][FVD96] or health[KKR06][WCW$^+$87].

Retail is also another area of interest, with empirical evidence on forecasting. Retailers grown from small shops to chain stores, with hundreds of locations. The challenge is not only forecasting in a store level, but also in chain and market levels[MKF18]. Forecasting grant is used to support decision making in strategic, tactical and operational plannings. The decisions will be key in distribution network, new store locations, position or even competitors performance. Retailers also have the need to forecast total sales and product demand. In sections 2.2 and 2.3, both concepts will be reviews with detail.

2.1 Cross Sectional and Time Series

Market researchers and business analysts when facing the problem of predicting sales have two approaches. One approach to prediction is to use cross-sectional data, working with one point in time and explanatory variables[FM04], such as, population, employment rate, personal income, precipitation, temperature, and population growth[Kon08]. In statistics and econometric, the methods of cross sectional are based on regression models[MFH16]. This type of cross-sectional analysis considers one point in time, with no particular order, in contrast to a time-series regression or longitudinal regression in which the variables are considered to be associated with a sequence of points in time.

In time series analysis past observations are ordered by a time period to estimate a future observation, using classical time series methods, such as ARIMA and Winter's[SLP15]. Claiming that the performance of classical methods could be improved using explanatory variables, researches proposed models that included it, such as SARIMAX (Seasonal Autoregressive Integrated Moving Average with External Variables) [AAF16].

2.2 Sales Forecasting

In this section, it is reviewed the concept of forecasting the total amount of sales a market, chain or store. The object of forecasting is measured in monetary units, instead of units of a product[MKF18]. The time bucket is an extended period like year, quarter or month, but particularly for store forecast, a day can better suit the need. Market, chain and store level forecasting have different objectives, but they have common issues, namely seasonality and trend.[MKF18]. Time series can model those issues, so they have been the most common approach to forecasting total sales[XDH08][AQS01][ABGM15].

Retail sales often exhibit strong seasonal variations. Historically, modeling and forecasting seasonal data is one of the major research efforts and many theoretical and heuristic methods have been developed in the last several decades[CZ03]. Traditionally time series models, such as ARIMA,an autoregressive integrated moving average, or exponential smoothing and its extension to include seasonality and trend(Holt-Winters), have shown good performance forecasting retail sales, especially in economic stability.[AQS01]. Traditional methods are linear and so they are tight to the conditions of linear models and cannot efficiently handle the nonlinear components in the data. Neural networks are a class of flexible nonlinear models that can discover patterns from the data[ZQ05]. The possibility of neural networks capturing nonlinear seasonality, and other nonlinear components, made researchers study which was the best solution using a neural network to forecast total sales: using the data directly or preprocessing it, modelling seasonality and trend.

Seasonality is a periodic and recurrent pattern caused by factors such as weather, holidays, repeating promotions, as well as the behavior of economic agents. Alon [AQS01], forecasting the retail sales of US, compared ANN with traditional time series models ARIMA and Holt-Winters, and also multivariate regression. ANN and regression, modeled seasonality using season dummy variables. The data was divided in two periods, each one was marked with instability, with high unemployment and interest rates. ANN outperformed other models and regression had the worst performance. Traditional models however, had good performance in the period that the conditions were more stable. It also states, that ANN can capture seasonality in the data, but is important to notice that the data was not raw and had preprocessing to model seasonality.

Aye forecasting South Africa retail sales[ABGM15], also modeled seasonality with dummy variables. It compared 9 full seasonal models with 17 dummy models. Full seasonal models data lacked performance. Also from the analysis, it states that was difficult to identify a specific model as the best for forecasting South Africa's aggregate retail sales. Some models are well suited for booms, while others are well suited for recessions, and this differs across forecast horizons.

A comparative study of linear and nonlinear models for aggregate retail sales[CZ03], forecasting US retail sales, modeled the seasonality in dummy variables, but also deseasonality and trigonometric functions. A deseasonality model is where the seasonal component has been removed and is called seasonal stationary. The linear approach modelled seasonality with ARIMA model and regression with dummy models or trigonometric. The nonlinear approach was represented by ANN. The overall best model was ANN with deseasonality data and trigonometric functions proved not to be useful in modeling retail sales seasonality. The direct data model, also had a worst performance than dummy season variables.

Zhang made a comparison of the performance of the forecast using 3 models of seasonality, raw data, dummy variables and deseasonality[ZQ05]. It states that ANN can't model seasonality directly and the best approach is also to deseasonality the data before processing.

Another component with empirical studies is trend. A time series with trend is considered to be non stationary and often needs to be made stationary before most modeling and forecasting processes take place[ZQ05]. In a non stationary time series the trend component changes over time and this can be a problem to perform a viable forecast. The literature in modeling trend is scarce, but removing it from the data had better performance then processing it raw[ZQ05].

To conclude, when forecast total sales with a high level of aggregation, seasonality and trend are two components that must be modeled to achieve a good result. Statistics models have in Holt-Winters and ARIMA two viable solutions and ANN looks promising, with seasonality and trend adjustment.

2.3 Product Demand Forecasting

Product demand forecasting is directly related to many operational decisions as pricing, space allocation, ordering and inventory management[MKF18]. The ability to estimate the demand of a product is critical to impulse the proper decision making in the retail sector being inventory management a prime candidate for perceived cost savings[EK04] and a pricing strategy must be optimized in order to maximize revenue[FOV12]. Given a decision making question, the product demand must be characterized in three dimensions (Fig. 2.1): the level in the product hierarchy, the position in the retail supply chain and the time granularity[MKF18].

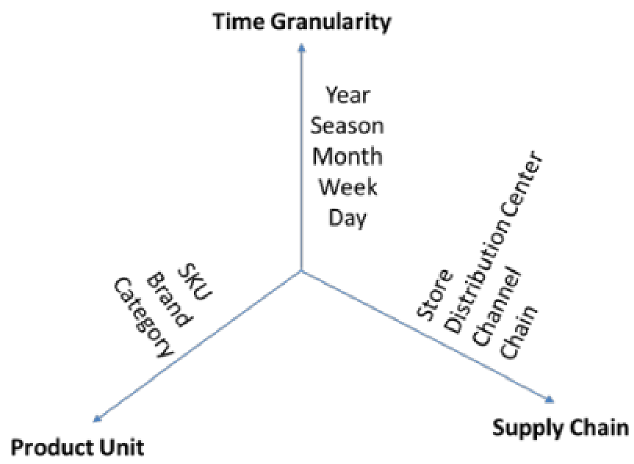

Figure 2.1: Multidimensional hierarchies in retail forecasting [MKF18]

Concerning product hierarchy, the approach is to aggregate by SKU (Stock Keeping Unit), brand or category. Retailers rely on demand forecasts at SKU level when making business decisions as marketing, production, inventory and finance[ASvWF09]. A SKU identify a single product and may identify the product in the whole chain store. Walmart, a large retail company, has roughly 5000 stores throughout the United States, meaning that approximately 1 billion unique forecasts must be generated in order to know the demand of all SKUs in the chain[Sea18]. A brand in a product category contains alternative SKUs with different packages types, sizes or colors. Brand level forecasts are important where there are cross-brand effects and promotions and ordering may be organized by brand [MKF18]. However, the initial forecast is made with a high level of aggregation, namely category forecasting, which contains a large number of SKUs with common characteristics[MKF18]. Using a larger model scope results in better forecast accuracy, by incorporating SKU data from multiple stores and multiple subcategories[ASvWF09].

The position on the supply chain highly impact the forecast. The supply chain consists of all intermediaries, from the manufacturer to the customer. To be able to forecast at this level, collaboration between parties is desired, which can be divided into two main categories: vertical. which include collaboration with customers, chain store and with suppliers; and horizontal, which include collaboration with competitors and with non-competitors[Bar04]. Theoretically, The collaborative forecast, would suppress the bullwhip effect. The bullwhip effect refers to increasing swings in inventory in response to shifts in customer demand as one moves further up the supply chain, leading to over inventory, stocks-outs and loss profit[Cha13].

Different time granularaties are need for different management decisions, from operational to strategic. From operational point of view, a fresh food store may need a forecast daily and a fashion store may have a need to estimate the demand to the whole season[MKF18].

2.4 Influential Drivers of Product Demand

Many factors influence the demand of a product. Some factors are in control of the retailers such as promotions, pricing and secondary effects like cannibalization or substitute products. Others not so much, but knowing their conditions is key. This is the case of seasonality, special days or weather[MKF18].

2.4.1 Weather

The supply and the demand of a product is affected by weather conditions[MDMFPL10], for instance temperature, humidity, snow fall, and, especially sunlight[LLLW11]. The consumer behaviour is affected by is mood and a higher temperature positively affect it[Par10]. Costumers expose to sunlight, natural or artificial, have a more consuming driven behaviour, willing to pay extra for a variety of products[LLLW11]. But this demand is not linear, as the temperature raises the consumption of soft drinks also increases, but when it reach extreme heat, the consumer adjust demand by opting for low sugar drinks or straight water[MKF18].

A challenge in use weather variables in forecasting is that are a wide range of weather conditions, such as temperature, wind, sunshine duration or precipitation. This must be forecast along with operational variables like prices or promotions and may cause retail not inclined to the extra effort[MKF18]. Nevertheless, weather can be used as a predictor variable, as example, forecasting beer demand, the highest daily temperature is a good estimator.[BF09]

2.4.2 Calendar Events

Since business activities and consumer behavior patterns may be greatly affected by holidays, the demand may vary substantially. Such effects are referred to as holiday effects[SLP15]. Demand can be also affected by special events such as festivals or sports. Specific calendar events can be modeled as seasonality, like Christmas. Alternately, Easter does not have a fixed date and cannot be treated as seasonality[MKF18].

Calendar events that are affected with a considerable increased demand, can be treated as a predictor variable. Events can be modeled for a single day, like Christmas or Easter, or a longer period, namely school vacations[AAF16]. As example, in New Year's Eve the beverages demand is marked with a huge spike and can be modeled as a dummy variable[BF09].

2.4.3 Seasonality

Product demand contains multiple seasonal cycles of different lengths,that can lead to inventory holding or stock outs[EHVW14]. Fashion retail have 4 main seasons, for instance spring,summer,winter and autumn, however, weekly and weekend is another type of seasonality. typically the paycheck is monthly, but can be weekly or biweekly, generating the correspondent seasonality. For this reason, models used in forecasting, where demand have strong seasonality must be able to handle multiple seasonal patterns[MKF18].

2.4.4 Promotions and Marketing

Promotions deeply affected the demand, are temporary and are "call to action". To benefit from the promotion the customer must acquire the product within a time frame[MP10]. In cooperation with price reduction, marketing with advertise can influence the costumer inside or outside the store[MKF18]. Promotion and marketing can affect the sale on a product, but also the sales of others products, this is called cannibalization[MKF18] and is strongly associated with brand switching[MP10].

Forecasting models should take in account the regular price and the discount price. Discount price may have various interpretations. Promotional price can be defined as the discount associated with a advertise or promotion campaign and discount price as the reduction that a product get after a few days in the shelf[AAF16]. A common use case is to use the relative discount, where a higher value means a bigger discount. This can be miss leading, when the discount is translated into dollar terms, the higher the price of the promoted product, the greater the amount of the dollar savings[CML98].

2.4.5 Stock-outs

When the demand of a product exceeds the inventory, a stock-out can occurs. This leads to an inaccurate demand estimation, because the focal product does not reflect the customer demand and substitutes product, being a viable option, may increase sales[MKF18]. The accuracy of the forecast directly contributes to higher profits by reducing stock-out situations[ASvWF09]

2.4.6 Intermittence

Also known as sporadic demand, intermittent demand, comes about when a product experiences several periods of zero demand. Often in these situations, when demand occurs it is small, and sometimes highly variable in size. This demand is hard to predict, and errors in prediction may be costly in terms of obsolescent stock or unmet demand[SB05].

Intermittence is hard to map, because a situation of zero demand, can be a result of another factor like a stock-out or seasonality[MKF18].

Current practice, where it uses statistical forecasting, favors exponential smoothing. An alternative specifically designed to deal with these types of patterns is Croston's. method[SB05].

2.5 Product demand forecasting applications

The improvement of models for forecasting product demand has been object of much effort for forecasting researchers and practitioners. The best model, is the one that fit the problem, with the easiest implementation and interpretation. In essence, the models can be split in 3 families: univariate, econometric and nonlinear[MKF18]. Univariate methods range from simple time series methods such moving average or exponential smoothing, to Box-Jenkins ARIMA and Holt-Winters[DVDVC+14] and state space[RSR15]. However, univariate models do not have in consideration external factors like promotions or pricing. Therefore, univariate models should be used forecasting on a higher aggregation or for products with low promotion or price elasticity.[MKF18]

Linear regression is feasible for large scale product level forecasting problems because is easy, simple and fast to fit. Econometric models use a joint approach of linear regression and exogenous variables such us seasonality, calendar events, weather conditions, price, and promotion features[MKF18].

PromoCast is a promotion event forecasting model and it uses a static cross-sectional regression analysis of SKU-store sales under a variety of promotion conditions, with store and chain specific historical performance information[CBL+08]. Forecasting the daily sales of banana, Arunraj developed a seasonal autoregressive integrated moving average with external variables (SARIMAX)[AAF16]. It was used exogenous variables to model seasonality, holydays and promotions.

Compared to the linear regression models, nonlinear methods allow arbitrary nonlinear approximation functions learned directly from the data and improves the potential to provide more accurate forecasts[MKF18]. Nonlinear methods include traditional nonlinear regressions, non or semi parametric regressions, and fuzzy and machine learning algorithms.

Machine learning is on trend, so recent studies emerged. A comparative study of Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function Network (RBFN), has shown that SVM had the best performance and it used holidays and the sum of all prices as predictive variables[SD17]. The implementation of a neuro fuzzy model using unit sales price, product quality, customer satisfaction, promotions, holidays and special days[EÖK09].

2.6 Summary

As discussed, forecasting is crucial in a company's strategic, tactical or operational planning. Total Sales and demand are representative of patterns of consumption, being sales crucial in strategic and tactical and demand on operational and tactical. Forecast techniques traditionally consisted of statistical models, but machine learning have been a new contender.

Chapter 3

Forecasting methods for time series

In this chapter, we introduce the models that will be implemented to forecast the demand of fresh fish. In section 3.1 we discussed statistical methods in general, section 3.1.1 the SARIMA method is explained, section 3.1.2 refers to Holt winter's seasonal method, section 3.1.3 explain the components of the model and section 3.2 is dedicated to the neural networks.

3.1 Statistical forecasting methods

The statistical methods implementation in this study are univariate. Univariate methods use only a single variable as input and most likely, the variable of input will be the output as well. Depending on the problem in hand, univariate forecasting methods can be extremely simple and surprisingly effective if used with the proper data. In our problem, we want to estimate the weight sold for a fish product 2 days ahead, so the input will be the daily sales till the present day. Concerning data complexity, this is very useful, because with only a single variable it's possible to develop a model that don't lack the accuracy of more sophisticated models, that use several inputs. The univariate methods explored can handle seasonality and trend, some may handle multiple seasonality, that is characteristic of retail data.

3.1.1 SARIMA model

This model is a improvement to an autoregressive integrated moving average model (ARIMA), but have the advantage that can handle seasonality. Before introduce the model in important to discuss the concepts **differencing** and **stationary**.

- A **stationary** time series is one whose properties do not depend on the time at which the series is observed. Thus, time series with trends is not stationary and will affect the value of the time series at different times. On the other hand, a white noise series is stationary — it does not matter when you observe it, it should look much the same at any point in time.

- **Differencing** is one way to make a non-stationary time series stationary — compute the differences between consecutive observations. Differencing can help stabilise the mean of a time series by removing changes in the level of a time series, and therefore eliminating or reducing trend and seasonality.[Hyn]

A non-seasonal ARIMA model is the combination of autoregression and a moving average model. In an autoregression model(AR), the variable of interest is forecasted using a linear combination of past values of the variable and the term autoregression indicates that it is a regression of the variable against itself. This model normally is restricted to stationary data. Rather than using past values of the forecast variable in a regression, a moving average(MA) model uses past forecast errors in a regression-like model. The integrated(I) is the initial step of the model, where is calculated the degree of Differencing. An ARIMA model can be described as ARIMA(p,q,d), where:

- p = order of the autoregressive part;
- d = degree of first differencing involved;
- q = order of the moving average part.

As said before, SARIMA or also referred as seasonal ARIMA, can handle seasonality and can described as:

- Non seasonal: (p,q,d)
- Seasonal: (P,Q,D)m
- SArima (p,q,d) . (P,Q,D)m

The seasonal part of the model use similar processes, where m = number of periods per season.

3.1.2 Exponential Smoothing Methods

The simplest of the exponentially smoothing methods is simple exponential smoothing(**SES**) and is weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight. This framework generates reliable forecasts quickly and for a wide range of time series, which is a great advantage and of major importance to applications in industry. The **SES** is suitable for forecasting data with no clear trend or seasonal pattern and relies on calculate the level of the series, in a certain point. In 1957, Holt[Hol57] extended simple exponential smoothing to allow the forecasting of data with a trend. Splitting the series in two components, level and trend, each one is calculated for each point of the series and then added to each other resulting on the forecasted value. Forecasts produce by these methods tend to over forecast, especially for longer forecast horizons, so Gardner and McKenzie[GM85] introduced a parameter that “dampens” the trend to a flat line some time in the future. In 1960, Holt and Winters proposed the Holt-Winters seasonal method[Win60], that add another component to the formula, the seasonality component that have 2 variations to be calculated. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

3.1.3 Tbats Model

Tbats is a recent model from 2011[dLHS11], and the name are acronyms for key features of the models: Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components,it can handle multiple seasonality, where each seasonality is modeled by a trigonometric representation based on Fourier series.

3.2 Artificial Neural Networks

Artificial Neural Networks (ANN) are distributed systems based on the nervous system and are composed of a set of artificial neurons, constituting processing units. Each artificial neuron has a set of input connections, to receive input values either from an input attribute vector or from other neurons. Each input connection has a weight value associated, simulating the synapses in the nervous system. The network weight values are defined by a learning algorithm. A neuron defines its output value by using an activation function to the weighted sum of its inputs. This output value is sent to the ANN output or to other artificial neurons. Figure 3.1 demonstrates a ANN structure with 4 inputs, a hidden layer and a output layer.

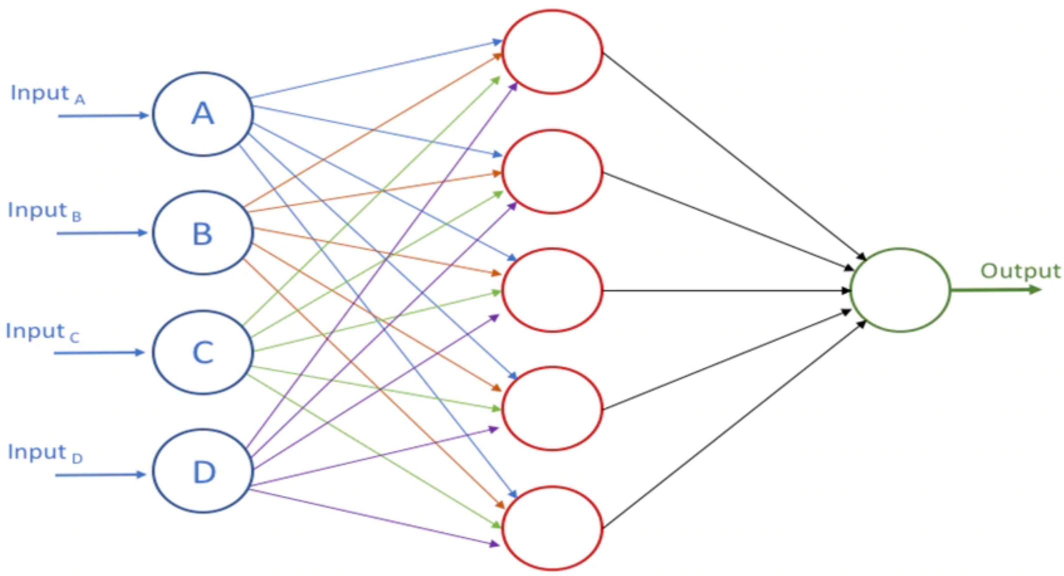

Figure 3.1: Diagram of the architecture of a neural network.

3.2.1 Long Short Term Memory

A Recurrent neural network(RNN) is a artificial neural network designed to recognize patterns in sequences of data, such as numerical times series data emanating from sensors, stock markets and government agencies. What differentiates RNNs from other neural networks is that they take time and sequence into account, forming a sequential of information. That sequential information is preserved in the recurrent network's hidden state, which manages to span many time steps as it cascades forward to affect the processing of each new example. It is finding correlations between events separated by many moments, and these correlations are called "long-term dependencies", because an event downstream in time depends upon, and is a function of, one or more events that came before. One way to think about RNNs is this: they are a way to share weights over time.

A Long Short Term Memory(LSTM) is a new type of RNN, that contains information outside the normal flow of the recurrent network in a gated cell. Information can be stored in, written to, or read from a cell, much like data in a computer's memory. The cell makes decisions about what to store, and when to allow reads, writes and erasures, via gates that open and close. Those gates act on the signals they receive, and similar to the neural network's nodes, they block or pass on information based on its strength and import, which they filter with their own sets of weights. Those weights, like the weights that modulate input and hidden states, are adjusted via the recurrent networks learning process. That is, the cells learn when to allow data to enter, leave or be deleted through the iterative process of making guesses.

3.3 Conclusions

In this chapter, we introduce the methods that will be implemented to forecast the demand of fresh fish. From the methods introduce, three are statistical models, where SARIMA ans HOlt Winter's seasonal methods are considered historical and an inspiration for recent methods like TBATS. We also explained the architecture of the artificial neural networks and the main concepts of a LStm. In the next chapter, we present the problem description and the exploration of the transactional data presented by the retail partner.

Chapter 4

Problem description and data exploration

This chapter describes the reasons of the forecasting needs and the transactional data. Section 4.1 describe the fish department of the retail partner and the forecasting need, Section 4.2 describes the relation between client card and overall sales, section 4.3 indicates the forecast goal and from section 4.4 till section 4.9 we explore the transactional dataset narrowing the level of aggregation.

4.1 Description of the forecasting need

The retail partner has 270 stores across Portugal and his fish department has a total of 107 fresh fish products, with 2 different supply chains, where 89 the source is **wild caught** and 18 from **aquaculture**. The orders from the wild caught supply chain are acquired in the providers early in the morning, when the boats arrive from the fishing and the outcome can not be enough to supply the demand, leading to a stockout. The supply chain is closed on Monday, so no orders are schedule to this day. The products provided are seasonal, mostly to ensure the preservation of the wild life. Products acquired in the aquaculture supply chain, being a industrialized source, can be fairly supply the demand.

Fish products are fast moving moving products with a high level of deterioration. The consumption of spoiled fish can be harmful and the orders for fish products must be made daily, so the offer available is the most fresh desirable. Forecasting the demand of hundreds of products can be a pain full task for management, specially, when the product can't be stored for a long period of time. To prevent the spoil, the orders are schedule daily, 2 days prior to the acquisition. This period ensure enough time to organize the acquisition in the providers and the transportation to the retail facilities. The estimation of the daily demand is directly related to the volume of the orders and the precise estimation will benefit by reducing waste, prevent stockouts and increase profitability.

As summary, the deterioration level of fish products leads to a need of daily forecast, but the next day is a brief window to organize the necessary process of the supply demand, so a 2 day ahead horizon is the goal of forecast.

4.2 Client card and overall demand estimation

The retail partner provides to the customer a loyalty program that provides direct discount on the supplied products. As part of the program, the customer is presented with a card to identifies him as a participant in a loyalty program. The transactional data of the loyalty program, was made available for this study, but this data does not contain the overall demand, because sales out of the loyalty program are not part of the program. The retail partner also provided that 85% of the sales result from the loyalty program, so is reasonable to conclude that the values used to forecast the demand of fish products in this study, are also 85% of the true demand.

4.3 Forecast Goal

The goal is to build a model of machine learning using neural networks, that forecast the value of a fresh fish product in the 2 day ahead horizon. To validate the forecast of the machine learning model, statistical forecasting models are also implemented.

The retail partner provided a dataset containing the demand for all the products from 3 stores, that are located in in Pedrouços, Sintra e Alcabideche. In the next section, we present the process taken in the exploration of the dataset.

4.4 Transactional Data

The first step towards building a forecasting model was exploring the transactional dataset provided by the retail partner. This dataset contains 4 tables: transactions, products,clients and locations. The transactions table contains over 100 millions transactions collected in 3 stores of the retail partner that are located in Pedrouços, Sintra e Alcabideche.

4.4.1 Reducing the dimensionality of the data

By reducing the information before the exploration, we benefit from focus our efforts in comprehending the most valuable information at our disposal.

The transactional dataset contains information for all the groceries sold by the retail partner and contains 148.568 products and 106.713.565 transactions. Since the goal is to forecast the quantity of fish sell sold, products and consequent transactions not regarding it were filtered from the dataset. After the filter applied, remained 107 SKUs associated with 407.492 transactions.

During the literature review, was not found information about the client profile as a influential driver of product demand, so the **clients** table was also removed from the preprocessing of data.

The final transactional dataset (fig. 4.1) contains transactions, clients and locations tables. At a first glance, the most valuable information from the database seems the following:

- **location_cd**: store the transaction take place.
- **time_key**: the moment when the transaction was registered.
- **sku**: identifies the product in the transaction.
- **quantity**: total or weight acquired by the customer.
- **specie**: the specie of the fish product, can contain multiple skus.

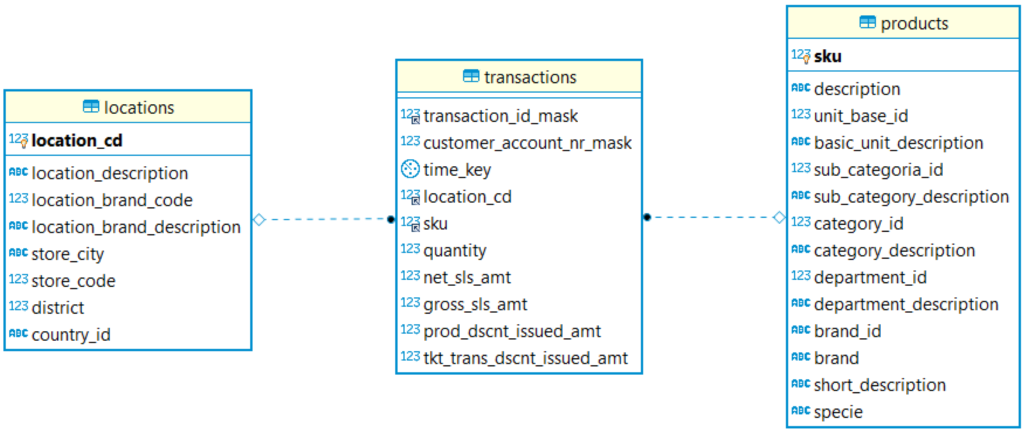

Figure 4.1: UML of the transactional dataset.

4.5 Data exploration

After obtained the information regarding fresh fish information, our efforts were redirected to improve our knowledge how the fish market behaves, so the starting points was a broad view of the market 4.6, next the data is desegregated to store level resulting in the most relevant species that identifies the demand 4.7. To conclude the exploration, we select 3 products of different species that will be applied the forecasting models 4.9.

4.6 Market level exploration

The first contact of the exploration, was a wide approach to analyse the demand behaviour in general, obtaining an overall impression of the fresh fish market.

Analysing sales from 1 January of 2017 till 30 September of 2019 (Fig. 4.2), it has a clear outlier in July of 2018, that coincides with the traditional festivals in Portugal and Sardine is a delicacy enjoyed in the festivals, so may impact the sales. The annual seasonality is obvious, with the sales slowing increasing till a spike around July and then fall off till the end of the year. A trend analysis is not so obvious, maybe exists but is residual and not possible to identify with this visualization.

Retail sales should have strong weekly seasonality, and like trend is not very obvious, but to reflect the importance will explored in section 4.6.1.

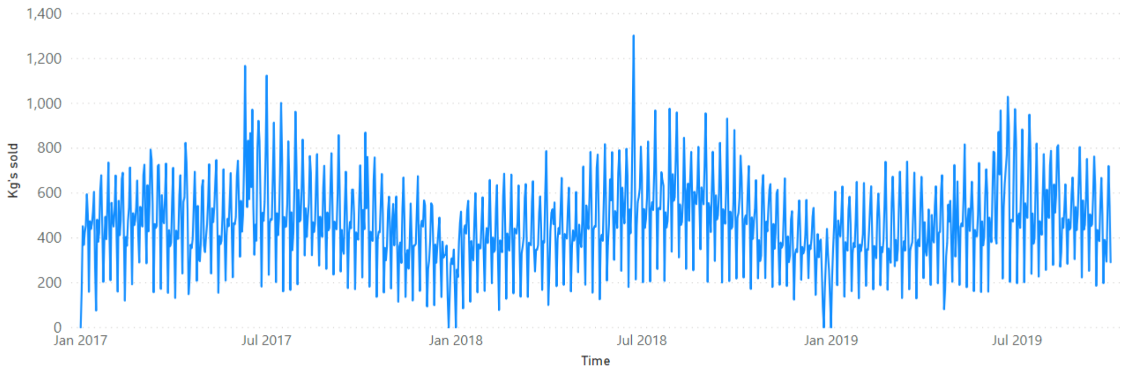

Figure 4.2: Demonstration of the fresh fish sales, for the 3 stores combined .

4.6.1 Weekly seasonality exploration

The visualization clear demonstrate the weekly seasonality of the demand, with the lowest point on Monday and the peak of sales on Saturday. The weekly seasonality is very important, because is the base to a forecasting model with daily aggregation. The existence of it it's a great starting point when it comes the times of implementing the models. Unfortunately, the values from Tuesday to Thursday are a sequence of values that do not differ much. Of course, here the aggregation is in a higher level, but it is a indicator the weekly demand behaves.

In 4.6, the trend was not obvious so a further exploration will be evaluated in section 4.6.2.

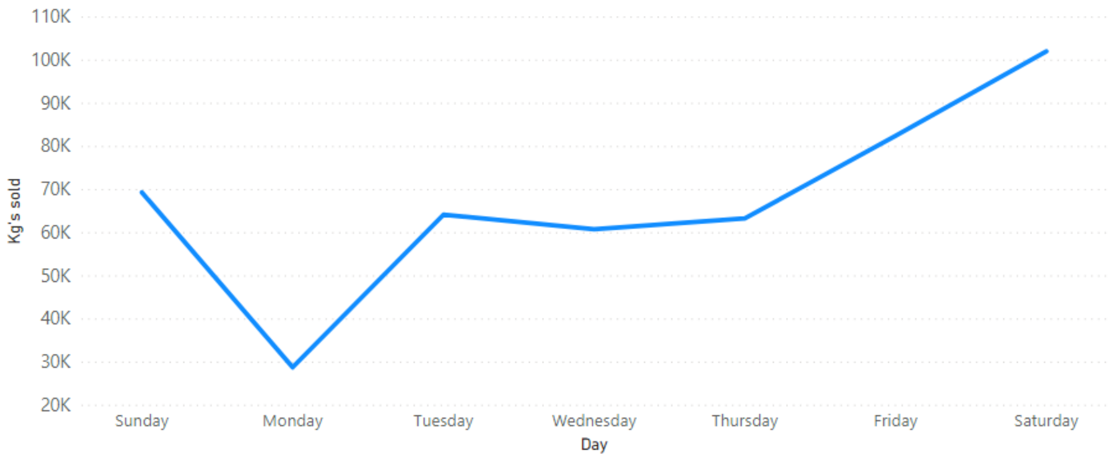

Figure 4.3: Daily sales of the first 10 weeks of the year.

4.6.2 Trend exploration

The yearly demand grants a clear visualization if exists a growing or diminishing trend, or in this case, if does not exists at all. Analysing 2017 and 2018, 2017 seems to have a higher volume of sales, that could mean a trend decrease, but after July the opposite happens with 2018 taking the upper hand on sales. Also, in March, 2017 has the top sales followed with 2019, instead of 2018. From this analysis, we can conclude that at least the overall sales of fresh fish does not contain the trend component, but does not mean when lowering the aggregation to the product level, it does not exist either. But before the product level aggregation, each store have to be explored, in a go deeper in the staircase approach. So, we explore the sales for the stores in section 4.7.

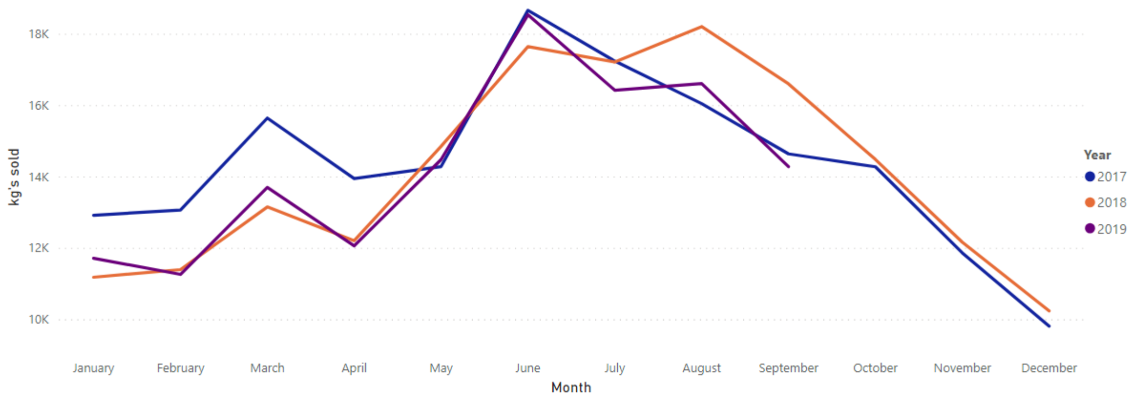

Figure 4.4: Visualization of the demand by year and month.

4.7 Store level exploration

The analysis in a store level grants information of how each store performs compared to the overall market exploration. But as starting point, was important to obtain the weight of each store in the transactional data 4.5, on the left side of the figure we have the code name of each store: 3, 246 and 326. Lets call each store by his code name, as examples, **store 3**.

Analysing the volume of sales of store 326, the quantity in only one quarter of the total sales from store 3. Even with the combination of stores 246 and 236, is still three quarters of the demand of store 3. This is a problem, such weight could compromised the results of the exploration on section 4.6, since the behaviour is most of the store 3.

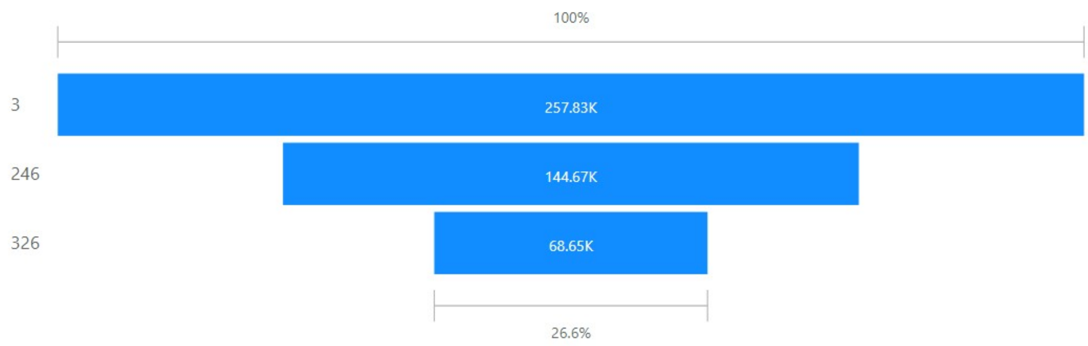

Figure 4.5: Total weight sold be each store.

However the analysis to the behaviour of each store 4.6 shows similar performance. Even the less alike, store 326, has a similar behaviour to the others stores. This behaviour has different values of magnitude, but can be estimated to be applied cross store.

All the stores use the same SKU for the products they have available for his costumers and the forecast horizon is a two day ahead for a given SKU, so the possibility to aggregate the sales of each store daily is reasonable. But the stores are geographically different and do not acquired the products in the same suppliers, so the forecast will be for an SKU on one store only. Since the store 3 has the most information on the data and his representative of his pairs, has been chosen for further exploration.

A retail store have multiple species available to the costumer, with different degrees of demand, that are splitted into multiple products that also have different degrees of demand. In section 4.8, we explore the favorites species, and the associated products, of the retail's costumer.

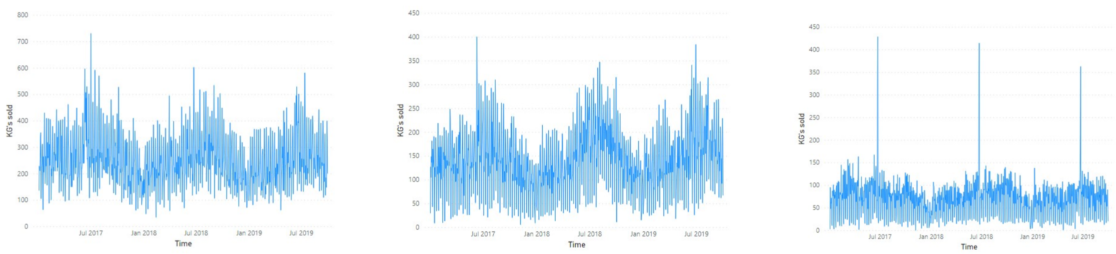

Figure 4.6: The performance of each store, on the left store 3, the middle store 246 and on the left store 326

4.8 Specie level exploration

The exploration at the specie level is a crucial process of the exploration, since we can narrow and group the SKUS we need to analyse on the product level exploration. The retail partner has available to the customer 107 different product, from 69 different species. Analysing the map 4.7 sales by specie, **dourada** clears stand outs and what looks like 25% of the sales. We choose to use the original Portuguese names, that are contained in the dataset, so if a research continues our work, it will benefit from the consistency on both sides.

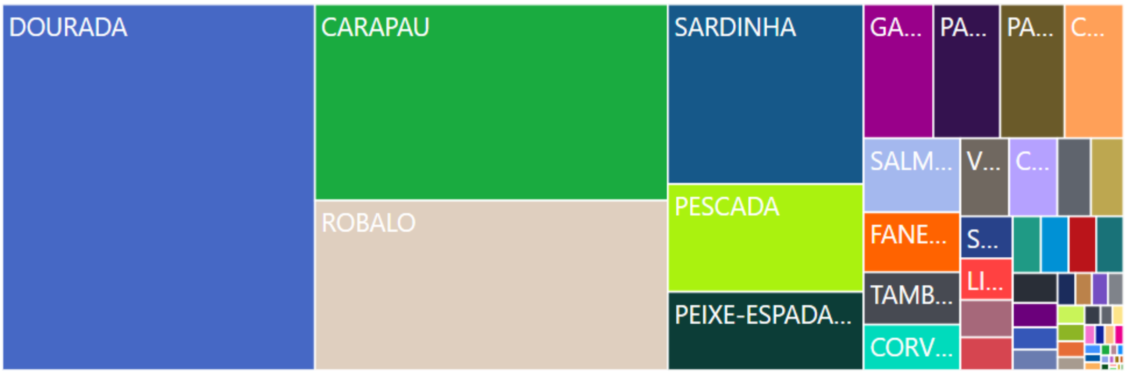

Figure 4.7: Tree-map of total weight sold of each specie in Continente Cascais.

Along with dourada, carapau and robalo are the species that the costumer most acquired in the stores of the retail partner, representing at least 50% of the total demand of store 3. An accurate forecast will obtain a greater impact on products with higher volume of sales and this species are more likely to contain those products.

We conclude the selection of the species to the product level exploration 4.9, but not before translating the select species to English terms, that will be used for the rest of the dissertation:

Dourada or gilt head bream.

Carapau or horse mackerel.

Robalo or sea bass.

4.9 Product level exploration

Finally we reached at the product level, where we select the SKUs to be used in the forecasting process. The analyse is done for each separately, on the product most representative of the specie is selected. We also take in consideration the number of zeros that the product has, is recognized that statistical models have issues dealing with zeros values.

4.9.1 Gilt head bream sales exploration

From the sales of gilt head products 4.8, "dourada média" stand outs as the product with the higher volume of demand and it sales are quite irregular and has a presence of an outlier around July 2018. However, the volume of sales is a strong factor and is the first product selected for forecast.

Analysing the pattern of sales, the annual seasonality does not look present in the data, but also looks that no signs of pattern on it. Regarding weekly seasonality, it seems that have some presence in the data, event if not follow the same pattern every week.

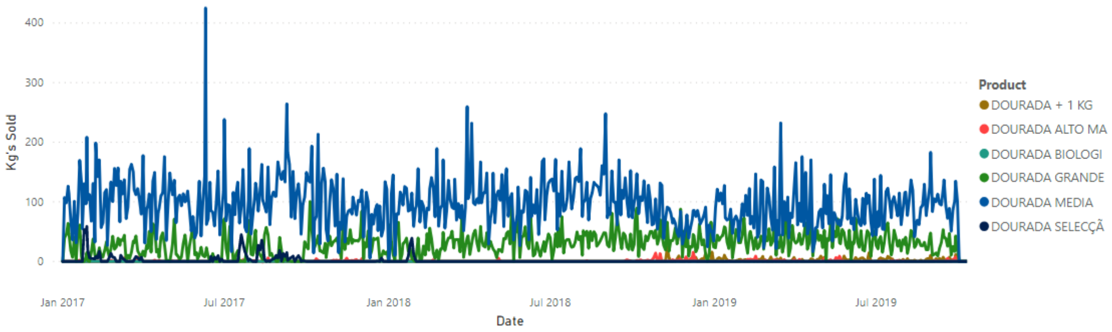

Figure 4.8: Chart of Gilt head bream sales from Cascais.

4.9.2 Sea bass sales exploration

The robalo sales have a similar performance of the dourada, but not with such a straightforward selection when picking one of the products. But a careful analyses show that "robalo grande" has imtermitence in the sales and contains multiple zero values and the selection drops on "robalo médio".

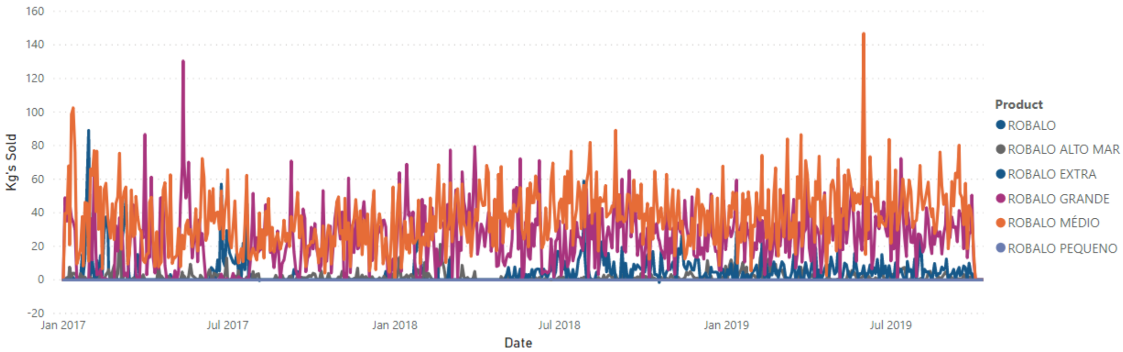

Figure 4.9: Chart of sea bass sales.

4.9.3 Horse mackerel sales exploration

Analysing the sales behaviour from the two primary products of horse mackerel, are the ones that selling pattern more resembles the pattern from the general sales, especially "carapau médio". The annual seasonality is clearly in 2017 and 2018, with the lower in January and the peak of sales around July. Is not clearly, but it seems that the data exhibit week seasonality. Forecasting models to forecast must take into account more then one seasonality category and must deal with annual and weekly seasonality.

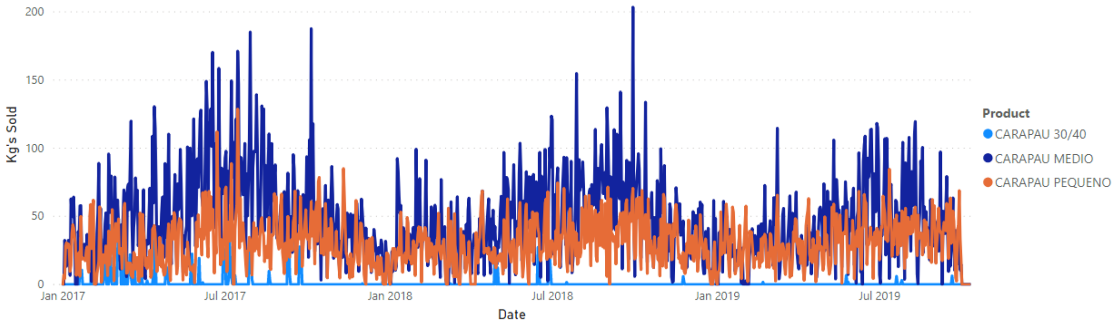

Figure 4.10: Chart of horse mackerel sales.

4.10 Conclusions of the data exploration

From the data exploration we conclude that the general sales have annual seasonality, but not all the selected products follow that pattern, in contrary to weekly seasonality component that all seems to imcorporate. No signs of trend was found in the data. We also discovered that even the transactional dataset contains three stores, store 3 contains more then 50% of the sales.

For an easier interpretation and since we only have one SKU per specie, when referring to each of the products, will be used instead the name of the specie in English for the rest of the dissertation, as demonstrate in the following list:

- Robalo médio (sea bass)
- Carapau Médio (horse mackerel)
- Dourada média (gilt head bream)

In the next chapter, the implementation of the forecasting models is discussed.

Chapter 5

Forecasting models implementation

In this chapter will be discussed the implementation of the forecasting models, that can be divided in statistical and machine learning. The statistical models are represented by SARIMA, Holt Winter's seasonal model, TBATS and Fourier seasonal modelling;machine learning is represented by a LSTM, a recurrent neural network.

In the first part, is presented the implementation of the statistical model and calculate the auto correlation function, to a better understanding of the series, finishing with the study of forecasting results.

In the second part, is explained the implementation of the LSTM model, the process of feature selection and the forecasting results of the neural network. The comparison between statistical and machine learning results is also done here.

During this chapter are presented multiple results to validate the models. The metrics displayed are always calculated using the testing data.

In the last section, we present the overall sales forecast, a dataset that contains only a period of 9 month that the previous models are applied to validate the results.

5.1 Forecasting with statistical models

By exploration the sales of the selected products, it was found that the time series contains weekly seasonality, so the models chosen to implement had to at least handle single seasonality. However, the sales of horse mackerel are composed with annual seasonality, so models that could handle multiple seasonalities were also implemented.

The statistical univariate methods were implemented in R Language, using the **forecast package**. The package provides multiple out of the box forecasting methods and is a good starting point in a forecasting project. The following models were implemented:

- **auto.arima** : Returns best ARIMA model according to either AIC, AICc or BIC value.

- **tbats**: Fits a TBATS model, it can handle multi-seasonality.
- **hw**: Fits a exponential smoothing model anf can only handle one seasonality.
- **fourier term's**: Fits an curve in Fourier terms and can be used to model seasonality. The transformation using different frequencies, makes possible the modeling of multiple seasonalities.

The sales of each selected species were forecasted using each model, this way providing a way of comparison and a more reliable end result. The 1003 daily observations were splitted in 730 training examples and the 273 remaining used for testing data.

The use of prior time steps to predict the next time step is called the window method. This window may shift one observation for each forecast, being called sliding window. The validation was implemented using an expanded sliding window, where all the observations prior the forecast point are used for the estimation. As referred in the problem analyses, the forecast horizon is 2 days prior the last observation. By using all the information till 2 days prior of the forecast, the model uses the most recent information in the outcome of the result. The study of the auto correlation function (ACF) can be used to know if a time series if suitable to use statistical models and is discussed in the next section.

5.1.1 Study of the Auto correlation

Just as correlation measures the extent of a linear relationship between two variables, auto correlation function(ACF) measures the linear relationship between lagged values of a time series. A time series without any auto correlation is completely unpredictable and may not be suitable for statistical forecasting methods.[Hyn]

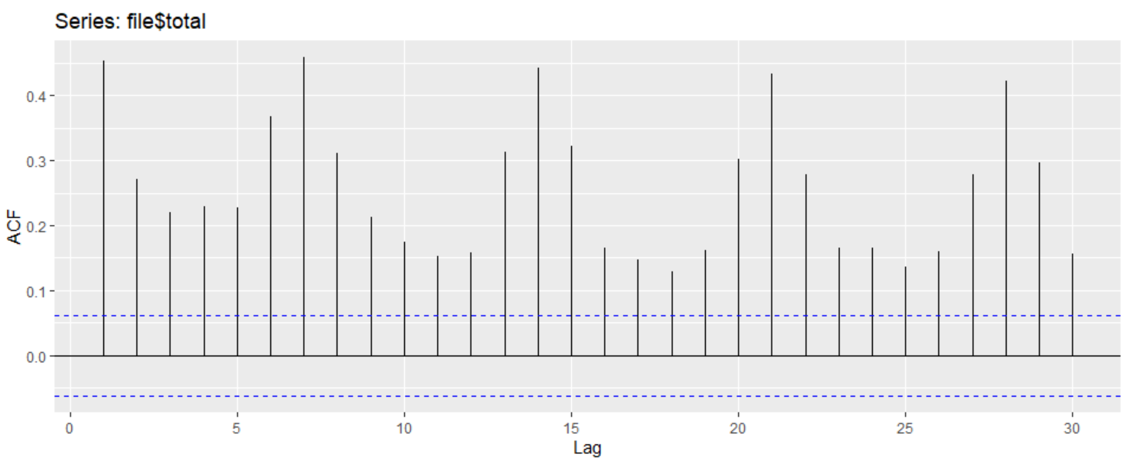

Figure 5.1: Auto correlation chart of the horse mackerel sales.

The auto correlation chart of horse mackerel sales (fig. 5.1) shows the auto correlation from lag 1 to 30. The dashed blue lines indicate whether the correlations are significantly different from zero. The highest point, lag 7, representing the same day in the previous week and continues to being notice as the lag increase. The rounded shaped between spikes demonstrates seasonality and slow decrease in the ACF as the lags increase is due to the trend, but is not very visible in the chart.

The ACF from sea bass sales are also representative of the Gilt head bream sales (fig. 5.2. The ACF demonstrate the spike at lag 7 and his pairs, but does not have the same degree on seasonality of the horse mackerel demand. The ACF confirms that the trend is residual in the sea bass sales.

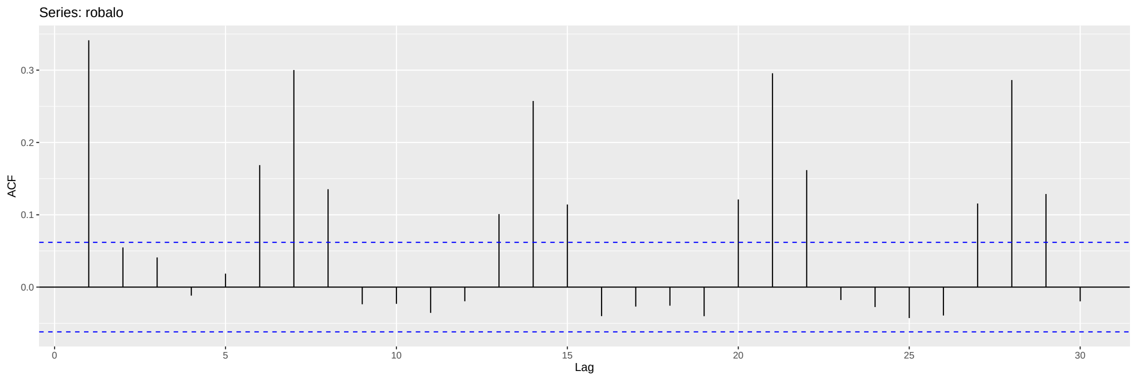

Figure 5.2: Auto correlation chart of the sea bass sales.

From the study of the auto correlation, we discover that the statistical methods are suitable to be used in the forecast of the chosen species. The results of the forecasting models are discussed in the next section 5.1.2 and the metrics displayed are calculated using the test dataset.

5.1.2 Forecasting results

The results obtained from the statistical models are presented on the table 5.1 and the metric chosen for evaluation was the NRMSE, because the differences on peaks have a higher weight, than differences on other values.

Table 5.1: Statistical models NRMSE values.

	SARIMA	HW	TBATS	Fourier's Terms
Horse Mackerel	0.154	0.154	0.149	0.149
Sea bass	0.149	0.146	0.139	0.139
Gilt head Bream	0.112	0.103	0.111	0.111

The models that model multiple seasonality, TBATS and Fourier's Terms, had the best performance, which was expected because a more complex model usually leads to higher accuracy. It looks strange that both had the same accuracy with 3 decimal numbers,but both use Fourier terms for capturing seasonality with ARIMA errors capturing other dynamics in the data and this may lead to the similarity in the results. Regarding the other models, SARIMA and HW had close performance, with the latter on top.

Multiple methods were implemented forecasting the demand, so in section 5.1.3 an approach to take advantage of it, is presented.

5.1.3 Models daily Evaluation by Ensemble Learning

To take advantage of the implementation of multiple models, the fitted values of the training dataset was splitted by week day to calculate the daily error of each model. The model that was most accurate in a given day, was chosen to forecast the same day of the test data. This is an ensemble learning approach, that uses the training data to choose the model to be used in the testing data by day.

Table 5.2: Ensemble process of the fitted values of the training dataset, on horse mackerel demand.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
SARIMA	0.148	**0.162**	0.170	**0.130**	0.118	0.184	**0.164**
Holt Winter's	0.151	0.176	0.171	0.139	0.115	0.180	0.166
TBATS	**0.141**	0.163	**0.162**	0.137	**0.114**	**0.175**	0.166
Fourier Term's	0.154	0.167	0.165	0.133	0.115	0.176	0.165
Ensemble	**0.141**	**0.162**	**0.162**	**0.130**	**0.130**	**0.175**	**0.175**

Analysing the results of the ensemble process 5.2, TBATS, on Sunday, was the most accurate forecasting the values of the training dataset, so consequently is chosen to forecast the Sunday's values of the testing period. The process continues for the remaining days, resulting in a ensemble learning method that group the most accurate models by day, forecasting the training period values.

The approach of a daily selection by the model's accuracy, may improve the result on the testing dataset. The comparison between the TBATS and Ensemble accuracy is presented on section 5.1.4.

5.1.4 Comparison of Ensemble Learning Method and TBATS

The best statistical model was TBATS along with the seasonal modeling with Fourier terms, but since they obtain the same result, only TBATS was picked for the comparison. Is a good practise,to compare the accuracy with a simpler model, to demonstrate the superior performance, so was the accuracy calculated using the average of the training data.

The TBATS model outperforms the average model in any of the selected species (table 5.3). The application of the ensemble method, only improve the accuracy of the horse mackerel sales and the accuracy on sea bass demand is lower when applied. This type of ensemble method, by selecting the most daily accurate model on the training data, did not prove that can be applied to all scenarios, but, when tested by specie can improve the forecast.

Table 5.3: Comparison of TBATS, Ensemble method and Average NRMSE values.

	TBATS	Ensemble	Average
Horse Mackerel	0.148	0.146	0.187
Sea Bass	0.139	0.140	0.172
Gilt Head Bream	0.111	0.111	0.117

To validate the results, we implemented multiple forecasting models. In section 5.2, we explained the process of implement a machine learning model. The model is a LSTM network with multiple features, where each feature is evaluated to understand the impact on the forecast.

5.2 Building an LSTM model

The LSTM was built using Keras that is a high-level neural networks API, written in Python. A LSTM network is an improvement to a RNN network, but contains a a memory unit that can maintain information in memory for long periods of time. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

A LSTM can handle multiple feature, in contrast to the methods described previous. Since the it can handle multiple features, the predictive of each feature must be evaluated to achieve a more accurate network. This features evaluates whether the features suggested in the literature are able to improve the model performance.

We also transformed the forecast in the differencing process. In a differencing process the objective variable is the difference between the current time step and a future value. The objective variable is the fluctuations in demand and not the true demand itself. Since the output is a difference, the input must be as well to achieve consistence in the data, so it's generated multiple differences features to lagged time steps.

5.2.1 Feature selection

The features can be categorized in differencing, date, price, meteorologic, lent and holidays, but difference has a special role within the group of features. As seen in table **??**, the target or the objective variable, is a input feature after 2 time steps. For this reason the differencing feature will be detach within the group of independent variables and referred as the main feature. The remaining are the secondaries variables.

Table 5.4: Target is a input feature after 2 time steps.

Timestep T(x)	Demand	Difference T(x) - T(x - 2)	Target T(x + 2) - T(x)
1	0	X	10
2	5	X	15
3	10	10	5
4	20	15	X
5	15	5	X

Multiple differencing features will be generated, grouped and evaluated the accuracy on the forecasting result. The groups that obtain the best result will be referred as the **main model**. The reason behind this, is that being in both sides of the equation, as input and output, is believed that it carries the highest weight in the forecasting result.

Then each secondary feature will be tested along with the it, to select the ones that most improves the accuracy of the main model. The one that most improves the main model, is considered the first configuration of the final model. This is considered the final of the first step.

In the second step, the remaining selected features are added gradually to the final model, from the one that performed the best till the one that performed the worst. If the model improves it becomes the new final model, the opposite makes the feature be discarded.

Since the the result of training a neural network is not precise and the result is different each iteration, in the first step each model will be trained using 20 iterations. In the second phase, will be 100 training iterations for a more consistence result. The models will be evaluated using RMSE.

For this process only one of the selected products will be used to evaluate the model building. The product under evaluation is the gild head bream, that had the lowest NRMSE in the statistical model forecasting. This strategy is taken, with the belief that if the product can achieve a better performance, the others species can benefit from that.

In the next section, is explained the process of the main or differentiation generation and selection.

5.2.1.1 Differencing features generation and selection

The use of differencing, modeling neural networks, may lead to higher accuracy. The forecasting result is not the actual sales value, but instead the difference between the present and the target day, as seen in table 5.5. Since the output is a difference, instead of modeling the week with the volume of sales in each day, the inputs were differences of the current time step for any other lagged day. This brings consistence between the input and the output value.

Table 5.5: 2 day ahead difference transformation.

	Total	Diff-2-day-ahead	Target Value	Target Diff
2018-01-01	10	#	5	-5
2018-01-02	15	#	0	-15
2018-01-03	5	5 - 10 = -5	#	#
2018-01-04	0	0 - 15 = -15	#	#

However, not only the present week has valuable information for the forecasting result and all lagged weeks can have a piece of information that could improve the accuracy. To accomplish this, the difference between the present and all previous days were generated, resulting in a number features equals to the number of previous days.

The first step consisted of evaluating how the accuracy behave by adding each week gradually as input. In the first iteration the model would only use 1 week, but in the second would use the 2 previous weeks. The model scores the best performance with the testing data, with 63 days or nine weeks 5.3. As such, We already know the time window required to obtain the best performance, but the model may still improve if we identify the features that contributes the most for the end result.

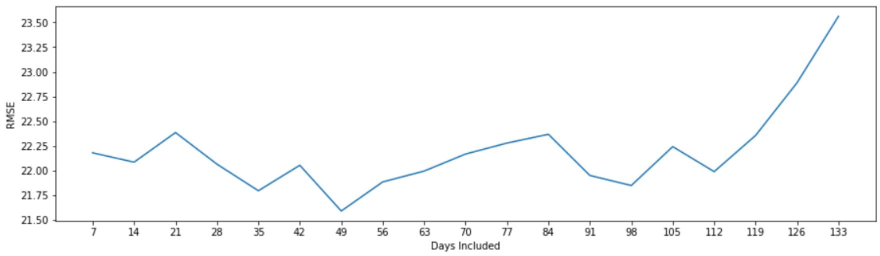

Figure 5.3: RMSE values by weekly inclusion of the main feature .

The 5 day difference to the present day obtained the best result 5.4, when the 9 weeks are included. For easier interpretation, this model will be referred as order 5 differencing, but in reality contain 5,12,19 and so on. The 5 order difference was the most powerful by far comparing to the other. This day, is the same of the target homologous day, but in the previous week. If the target is a Wednesday, this represents the last Wednesday and the inclusion of the last nine Wednesdays, produces a result close the inclusion of the last 49 days.

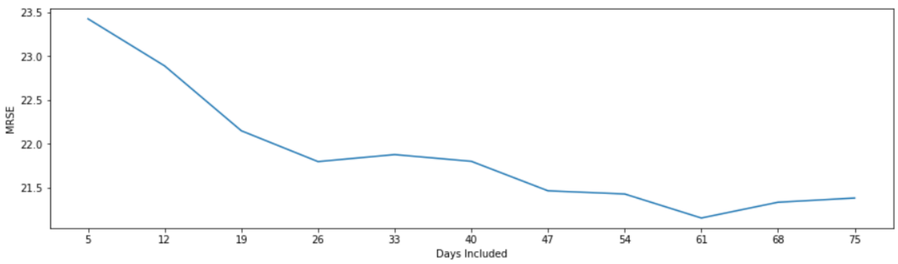

Figure 5.4: Order 5, the feature that obtained the best result of the differencing process.

The next step was to add gradually each week of the others differencing orders, but lock the order 5 differencing in the same week. Since the two previous tests obtain mixed results, the values of 7,8 and 9 weeks were used for the 5 order differencing. The model that obtained the lowest error contains information about the 9 previous weekdays of the target variable and the behave of the current week. The inclusion of information of lagged weeks, did rather then the present does not improve the performance.

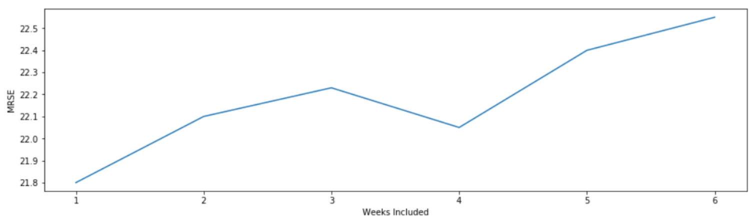

Figure 5.5: Best differing model, 9 weeks of the 5 order and 1 week of the others.

The dataset contains almost 3 years of data, so it is possible to include the difference of the current time step to the target value in the last year, representing a 362 difference, so it belongs to the order 5 series. Unfortunately, this impacted the accuracy negatively. Each feature of the same order has some correlation with each other (fig. 5.6) and the inclusion of multiple features with the same order is a trade off. For each feature of the same order that is added to the network, some correlation is added along with it, but the model also gains information about the lagged values, but then it reaches a point that is hard to estimate the correct weight to all this correlated features. At this point, the model starts to lose performance because is overwhelmed with correlated information.

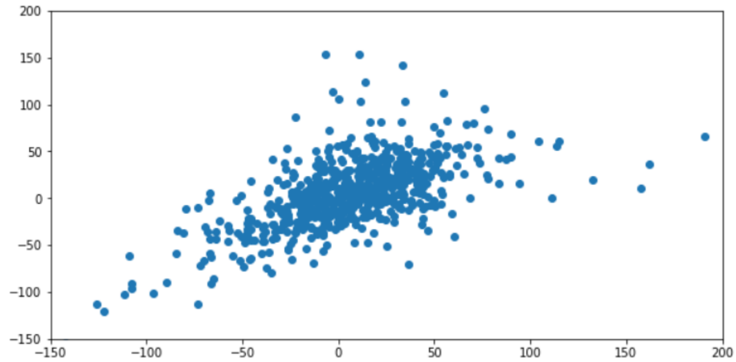

Figure 5.6: Scatter plot of of the differences of order 5 and 362, showing correlation with each other.

As conclusion, the main model will include the following differences as features:

- 1,2,3,4,6,7: representing the days in the week of the current time step, less the 5 day difference.
- 5,12,19,26,33, 40,47,54,61: the 9 previous weekdays of the target.

In the following sections, the evaluation of the secondary features are explained.

5.2.1.2 Holidays Features Modelling

The holidays were modelling using dummy variables, but with 2 different features types:

General holiday: this feature informs the model if the correspondent day is an holiday.

Specific holiday: exists 13 national holidays in Portugal, so for each was created a dummy variable, corresponding to 13 features.

We model the the holidays in this two ways, because the knowledge associated with tomorrow being an holiday, is not equal that being Christmas day. So the objective was that along with the knowledge of an general holiday represents, the weight of that specific holiday was also taken into account of the forecast.

To validate the predictability of each type of feature, 3 test cases were implemented:

- **Case 1**: the use only the general holiday.
- **Case 2**: the use of the 13 holidays dummies.
- **Case 3**: the use of both type together

The inclusion of holidays features leads to a substantial improvement in any of the type cases to the base model. The modelling of each holiday as a unique feature, has the most improvement, but used along the general holiday dummy seems that has a most stable result.

As conclusion, both holidays features has been selected to the second step.

Table 5.6: RMSE values of the test cases of the Holidays features.

	Average	Min	Max
Main	21.36	21.19	21.54
Case 1	21.03	20.69	21.33
Case 2	20.50	20.21	21.34
Case 3	20.46	20.23	20.60

5.2.1.3 Lent Feature

Lent is the six week period leading up to Easter and in Portugal is called the "Quaresma". It's one of the most important times of year for many Christians around the world and in particular Portugal, because 80% of the Portuguese population is Christian.

In Portugal, during this period, there is the common practise of the abstinence of meat and use fish as replacement. This event may lead to a higher volume of sales of fish products and to evaluate it was modelled as a dummy variable. Must take notice, that the period is the 40 days before Easter, that is a rotating holiday and so is this period.

The evaluation as predictable variable of the Lent period was perform with 2 test cases:

Case 1: the inclusion of the Lest along with the Base model.

Case 2: testing combined with the selected holidays features. The combined approach may model the holidays differently during the Lent.

The inclusion doesn't seem to impact the forecasting result on both cases. The reason why lent is not a predictable variable, is not the focus of the study, so we can only speculate the reasons:

- The lent does not impact the sales at all.
- Is a long period and the impact may be observed by the base model features.
- As a long period, may the impact exists, but not along all the duration of it. A prior study of the real impact duration may lead to better results.

Table 5.7: RMSE values of the test cases of the Lent feature.

	Average	Min	Max
Main	21.36	21.19	21.54
Case 1	21.41	21.17	22.04
Case 2	20.50	20.22	21.01

5.2.1.4 Price Feature

Any special technique was used to model the price feature and was it used raw. The price was calculated dividing the total amount of sales in a single day by the total weight. In the days with no sales, the price couldn't be calculated so the value of the previous day was used as input.

The price turn out to be a powerful predictor as seen in it's test case, lowering the error on 1.40 value from the base, so it was select to the second step of the feature selection.

Table 5.8: RMSE values of the test cases of the price feature.

	Average	Min	Max
Main	21.36	21.19	21.54
Price	19.96	19.64	20.63

5.2.1.5 Meteorologic Features

In the literature review it was possible to understand that multiple studies shown that meteorologic features can improve the forecast result, thus, the average temperature of the air and the total precipitation on a day were evaluated in 3 test cases:

- **Case 1:** Only the temperature is included.
- **Case 2:** Only the precipitation is included.
- **Case 3:** Both temperature and precipitation are included.

From the result of the test cases (table 5.9), the first conclusion was that the temperature impacted the accuracy negatively. Precipitation despite having a positive performance, the improvement in the accuracy is not enough to conclude at this time that has predictive power, but was select for the second step of feature selection for further investigation.

Table 5.9: RMSE values of the test cases of the meteorologic features.

	Average	Min	Max
Main	21.36	21.19	21.54
Case 1	21.81	21.57	22.15
Case 2	21.30	21.23	21.60
Case 3	21.70	21.57	22.03

5.2.1.6 Date features

Date was including in the model, but no modelling was implemented to improve the predictability of date features. The date features used were:

- WeekDay: Day of the week, from 1 to 7.
- Month: Month number in a year, from 1 to 12.
- WeekNumber: Week number in a year, from 1 to 53.
- Day: Day in a year, from 1 to 365.
- MonthDay: Day in a month, from 1 to 31, depending on the month.

The test case for the feature was inconclusive, because the error is lower, but it can't be directly related to the inclusion of the date features, since the result of the LSTM is not precise and it can fall in a range of values. However, date features were included in the second step of the feature selection for further investigation.

Table 5.10: RMSE values of the test cases of the date features.

	Average	Min	Max
Base	21.36	21.19	21.54
Date	21.29	20.92	21.72

5.2.1.7 Second step of feature selection

In this step, instead of testing each feature separately, they will be added gradually to the model, if the accuracy improves, that feature is selected to the final model, otherwise, the feature is discarded. The features under testing in the phase are:

- Holidays
- Date
- Precipitation

The price feature is not under evaluation, because it achieved the best accuracy and it's already part of the final model. In each test case, the "X" points which features are included in the training (table 5.11). The best performance was the combination of all the selected features, with an error of 18.20 and no feature is discarded in this step.

However, from the results there is no concrete proof that the date feature improves the performance, however, since is does not impact negatively it seems reasonable that is selected to the final model. Holidays and precipitation have similar performance, showing good predictive power.

Table 5.11: RMSE values for the test cases in the end of the 100 iterations.

	Main Model	Price	Holidays	Date	Precipitation	Average	Minimum	Maximum
Case 1	X	X				19.96	19.64	20.63
Case 2	X	X	X			19.55	18.65	21.36
Case 3	X	X	X	X		19.46	18.38	21.34
Case 4	X	X	X	X	X	19.28	18.20	20.32

5.2.2 Results of the feature selection process

Lent and temperature are discarded as predictive features. Anyway, further investigation is required, since they might be predictive, but the impact may is being absorbed by another feature.

The final machine learning model contains differencing, price, holidays,date and precipitation features, where price had the most impact in the forecast, followed by holidays and precipitation. In the next section, the results of the machine learning are presented and compared to the statistical modelling.

5.2.3 Comparison of the LSTM model with statistical models

The LSTM was build using the accuracy on forecasting the sales of gild head bream as the benchmark, but a good forecasting model should be versatile and performs as well when applied to another product. Comparing the most accurate statistical model for each specie and the LSTM (table 5.12, it shows that the LSTM was more accurate that any of the statistical models. The forecast results seems promising and machine learning has the advantage, that is a continuous process and new features can be included to improve the forecast. The advantage of the application of the statistical methods, is the simplicity of the implementation with the use of only variable, has almost the same accuracy of an LSTM with 38 variables.

In the next section, the best forecasting models for each product are presented.

Table 5.12: Comparison of the most accurate statistical model by specie and the LSTM model.

	Statistical	LSTM
Gilt head bream	0.103	0.088
Horse mackerel	0.149	0.148
Sea bass	0.139	0.133

5.3 Demonstration of the forecast result

To conclude the forecast results, the figures 5.7, 5.8, 5.12 are a demonstration for each product, of the forecasting result and the real sales value.

5.3.1 Gilt head bream result with LSTM

From analysing the forecast result, the model event with an irregular weekly seasonality, the model has enough accuracy modeling it. The outlier was detected, but not in the magnitude of his value, but just the detection can be seen as a positive sign.

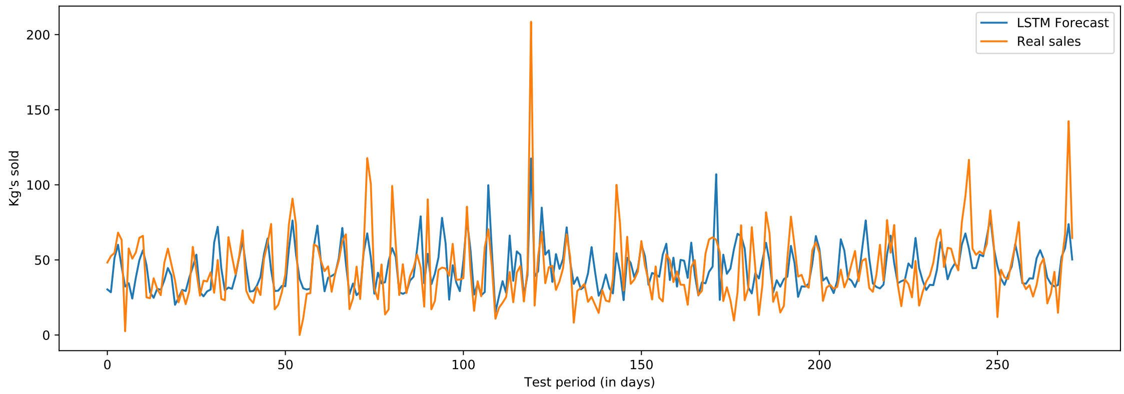

Figure 5.7: LSTM forecasting result of gilt head bream sales.

5.3.2 Sea bass result with LSTM

The LSTM is not very accurate in the start of the testing period, but after around day 75 the demand stabilizes and also the model results. The forecasting result does not seem very similar, but the sales of sea bass contain a high degree of irregularity, almost looking like white noise.

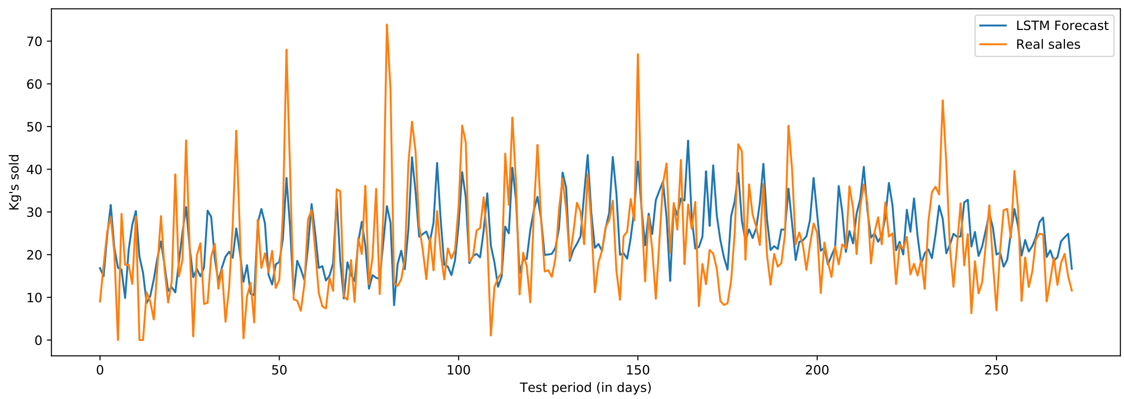

Figure 5.8: LSTM forecasting result of sea bass sales.

5.3.3 Horse mackerel result with LSTM

The LSTM had the highest error forecasting the demand of the horse mackerel. Like the other species, the sales are very irregular but the model had a satisfactory performance, capturing the outlier around day 70. Between days 150 and 200, the irregularity is clear and the model lacks some accuracy, but after that it follows the demand quite well.

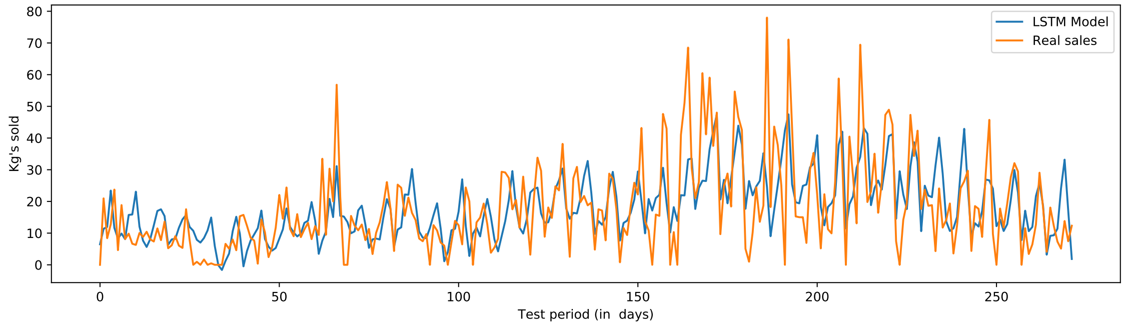

Figure 5.9: LSTM forecasting result of horse mackerel sales.

5.4 Forecast the overall sales, Benchmark with the retailer's method

The retail partner has also provided the overall sales of the selected products, but only for 10 months period - from 1 January until 22 of October of 2019. The dataset was provided to compare the results of the client card forecasting sales, with the result of the model that the retailer currently uses. Nevertheless, the card sales are a percentage of the overall sales and the direct comparison does not seem reasonable. Anyhow, the models to forecast the client card sales, were implemented to forecast the overall sales and compared to the retail's model performance. Along with the comparison with the retail's model, two additional observations can be drawn from this implementation.

First, The retail's model is used in production to forecast the demand of fresh fish products for more then 250 stores across Portugal, providing a reliable solution to schedule the orders. The implementation of the statistical models is straightforward using R language and the data is univariate, requiring only the demand modelled in a time series. A company in need of a reliable and low cost model, could gain from implementing in R language.

Second, Neural networks are very data depending, but the dataset only contains 252 observations and in the previous LSTM model the first 63 steps were used to calculate the first batch of differencing features. For this reason only 4 weeks, will be used to calculate the differencing. Anyway, there is left 224 observations where 88 are for testing, 30 for validation and remains only 106 steps for training. The accuracy might be impacted for the lack of information or the LSTM is powerful enough to model the demand. Also, the features used in the LSTM will be only differencing and price. The reduction of the number of variables happens because some become obsolete, being a one time event, but mostly, to study the performance of the LSTM with univariate type like model, but not discarding the price because it has significant predicting power.

In the next section, the results of the implemented models are presented along with the comparison with the retail's model.

5.4.1 Results forecasting the overall sales

Starting with the results of the statistical models, the top performance goes to TBATS and HW with similar accuracy. Forecasting horse mackerel and sea bass, HW was more accurate but TBATS was better forecasting the sales of gilt head bream. Sarima results during this dissertation, was not the most accurate compared to the other models, but no conclusion can be made since could be affected by the data used in this study.

The LSTM did not perform well forecasting the demand of horse mackerel. The model was built concerning the accuracy forecasting the sales of the gilt head bream, so there is the possibility that some modifications before applying to the horse mackerel data would improve the result.

On comparing with the retailer's model, the most gain is on the gild head bream with 0.34 improvement on the accuracy. Sea bass has a similar accuracy regarding the model and horse mackerel could not be improved from the retailer's accuracy.

We conclude this chapter, with the discussion and visualization of the forecasting results, in sections 5.4.1.1, 5.4.1.2 and 5.4.1.3

Table 5.13: NRMSE values of the overall sales.

	SARIMA	TBATS	HW	LSTM	Retail's Model
Horse Mackerel	0.184	0.175	0.168	0.189	0.148
Sea bass	0.143	0.136	0.133	0.123	0.128
Gilt Head Bream	0.141	0.116	0.118	0.104	0.138

5.4.1.1 Horse mackerel overall demand result

The model detected the downward trend in the demand for the test period, but has no success detecting the zeros sales values. The forecast seems accurate, but not detecting the break in sales around day 40.

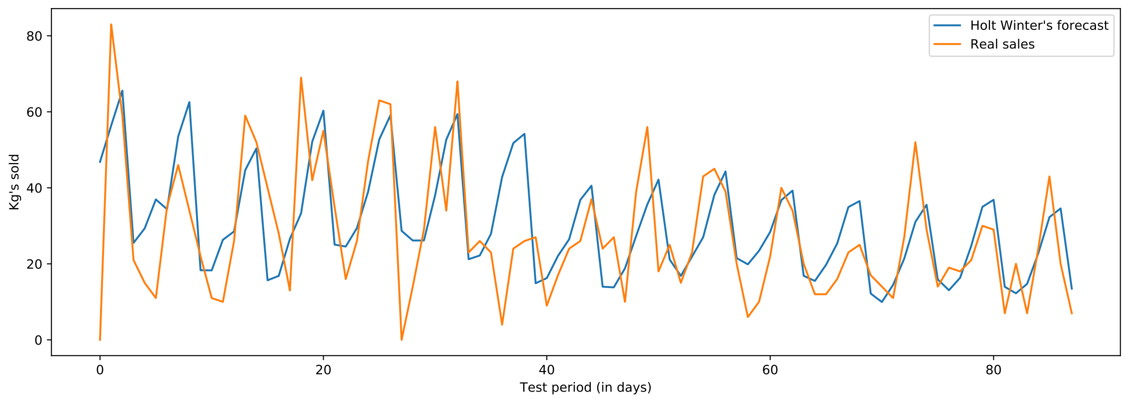

Figure 5.10: Holt Winter's forecasting of the horse mackerel overall demand.

5.4.1.2 Sea bass overall demand result

The LSTM model looks accurate forecasting the demand of sea bass, but especially in the 4 first weeks that the demand is very regular.

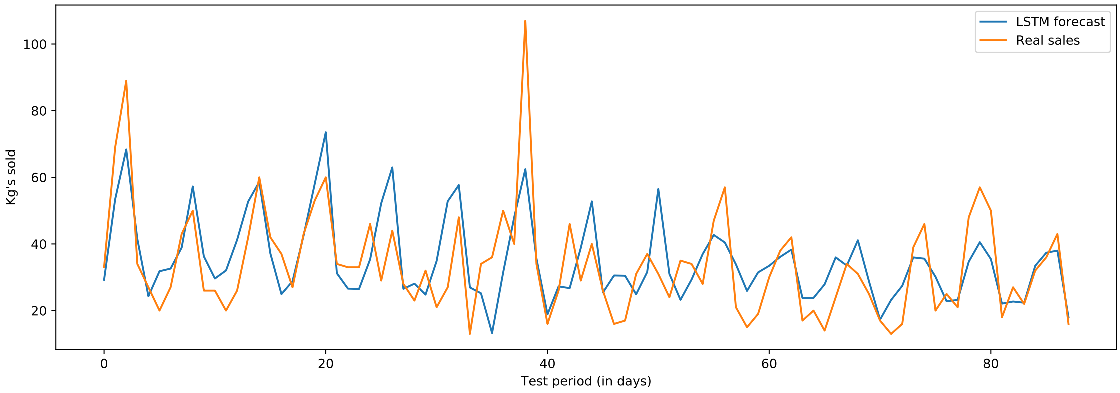

Figure 5.11: LSTM forecasting result of the sea bass overall demand.

5.4.1.3 Gilt head bream overall demand result

Gilt head bream has 2 outliers, that the it seems the LSTM could identifies the second, but not reached the same value of the true demand. Withdrawing both outliers, the forecast is accurate predicting the peaks and falls points of the demand.

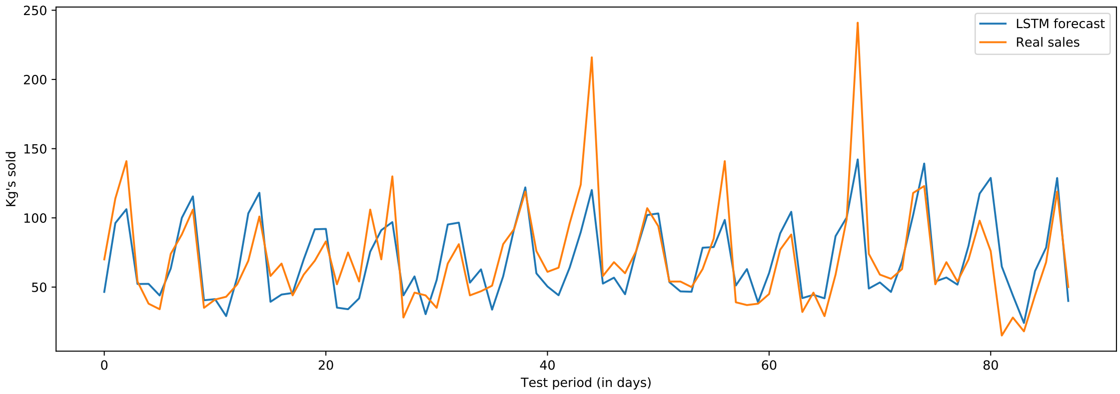

Figure 5.12: LSTM forecasting result of the gilt head bream demand.

Chapter 6

Conclusion

This chapter provides a summary of the study developed and the conclusions drawn. The conclusions regard the comparison of both type of models, the usability of the statistical model, why multivariate models are a doubtful advantage and the accuracy of the LSTM model with a simple network.

In this dissertation to forecast the demand of fresh fish products was implemented traditional statistical methods, such as SARIMA, Holt Winter's seasonal and Fourier seasonal model. This models provides reliable accuracy to the compared the the performance of the machine learning model, an LSTM network. Artificial neural networks have mixed results in the literature and this dissertation aims to be another contribute to unveil the mystery if they are suitable to model time series.

During this study, we reached the following conclusions:

a. Statistical modeling the demand of horse mackerel had the upper hand, but LSTM model was more accurate in gilt head bream and sea bass. This mixed at first can be interpreted as a negative result, because the lack of an objective result, but also can be seen as diversity, to be applied in different situations. The LSTM being a more complex model must be more accurate and from the results of this studies it seems. But the modelling and implementation is not as straightforward as the statistical models, with stable implementation packages with no need of manual tuning of the parameters. So, if we need a fast and reliable option we should opt for statistical models, but we have the time and also the need of a more accurate forecast, machine learning should be the option.

b. The usability of the statistical models is a big positive point that shouldn't be ignored, since the implementation in R language with the forecast package, is pretty straightforward for someone with programming background. The performance compared with the the Retail's model is very similar, which is used to forecast the demand for hundreds of stores. Not all the companies are the same and some companies may have the need of a more sophisticated model, but any company in need of a precise, but also low cost and maintenance forecasting model should take this option into consideration.

c. The most positive point of the LSTM model, is also is downside. The model is a multivariate, so it can be seen as a continuous process that any new variable found in research or practise can be added to improve the accuracy of the current model. This is very appealing, because the model is never obsolete, just have the need to refactor to include the new variable. The downside of this, is that being so variable dependent, not only the model must be maintained, but also all the sources of this variables. So the implementation of multivariate models, should only be applied in cases that the accuracy is the number one priority.

d. When forecasting the overall demand, only 2 variables were used modelling the LSTM, differencing and price. Also the period of training was only 106, but the network still presented a very accurate result comparing to the retail's model. LSTM is a recent model, much powerful then older networks architectures, that being a recurrent network is suitable to model time series. This is evidence, that past research of the application of neural networks to model time series is paying off.

Has a final statement, the results obtained were favorable and the LSTM obtained an accurate forecasting, that consolidate the application of neural networks to model time series.

6.1 Directions for future research

Forecast the demand is continuous process with always space for improvement, by the implementation of different methods or the inclusion of new features.

This dissertation evaluate the performance of LSTM forecasting the demand, but another recurrent network can be applied to forecast and evaluate the result. Gated recurrent units (GRU) has similar architecture, but uses gates instead of units as the base component. The application will benefit as a comparison of accuracy, but also as another method to forecast demand.

In the feature selection step, generating the differencing features, a experience was made to include the values of the past year, but the accuracy was impacted negatively. A more precise approach could benefit the forecast result, especially in products that have annual seasonality.

Finally, the models implemented forecast the demand of a single product separately and in terms of usability, was a benefit to forecast multiple products at the same time without the impacting the accuracy.

References

[AAF16] Nari Sivanandam Arunraj, Diane Ahrens, and Michael Fernandes. Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, 7(2):1–21, 2016.

[ABGM15] Goodness C. Aye, Mehmet Balcilar, Rangan Gupta, and Anandamayee Majumdar. Forecasting aggregate retail sales: The case of South Africa. *International Journal of Production Economics*, 160:66–79, 2015.

[AQS01] Ilan Alon, Min Qi, and Robert J. Sadowski. Forecasting aggregate retail sales: a comparison of articial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, 8(3):147–156, 2001.

[ASvWF09] Özden Gür Ali, Serpil Sayin, Tom van Woensel, and Jan Fransoo. SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 2009.

[Bar04] Mark Barratt. Understanding the meaning of collaboration in the supply chain. *Supply Chain Management*, 9(1):30–42, 2004.

[BF09] Danijel Bratina and Armand Faganel. Forecasting the Primary Demand for a Beer Brand Using Time Series Analysis. *Organizacija*, 41(3), 2009.

[CBL$^+$08] Lee G. Cooper, Penny Baron, Wayne Levy, Michael Swisher, and Paris Gogos. PromoCast$^{\mathrm{TM}}$: A New Forecasting Method for Promotion Planning. *Marketing Science*, 18(3):301–316, 2008.

[Cha13] Yu Fang Chao. The Bullwhip Effect in Supply Chain and Countermeasures. *Advanced Materials Research*, 711:799–804, 2013.

[CLC$^+$13] Carlos Cardoso, Helena Lourenço, Sara Costa, Susana Gonçalves, and Maria Leonor Nunes. Survey into the seafood consumption preferences and patterns in the portuguese population. Gender and regional variability. *Appetite*, 64:20–31, 2013.

[CML98] Shih Fen S. Chen, Kent B. Monroe, and Yung Chien Lou. The effects of framing price promotion messages on consumers' perceptions and purchase intentions. *Journal of Retailing*, 1998.

[CZ03] Ching Wu Chu and Guoqiang Peter Zhang. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, 2003.

[dLHS11] Alysha M. de Livera, Rob J. Hyndman, and Ralph D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 2011.

[DVDVC+14] Claudimar Pereira Da Veiga, Cássia Rita Pereira Da Veiga, Anderson Catapan, Ubiratã Tortato, and Wesley Vieira Da Silva. Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models. *WSEAS Transactions on Business and Economics*, 11(1):608–614, 2014.

[EHVW14] J. C.F. Ehrenthal, D. Honhon, and T. Van Woensel. Demand seasonality in retail inventory management. *European Journal of Operational Research*, 2014.

[EK04] A. H.C. Eaves and B. G. Kingsman. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4):431–437, 2004.

[EÖK09] Tuğba Efendigil, Semih Önüt, and Cengiz Kahraman. A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. *Expert Systems with Applications*, 2009.

[FM04] Edward W. Frees and Thomas W. Miller. Sales forecasting using longitudinal data models. *International Journal of Forecasting*, 2004.

[FOV12] Juan Carlos Ferrer, Diego Oyarzún, and Jorge Vera. Risk averse retail pricing with robust demand forecasting. *International Journal of Production Economics*, 2012.

[FVD96] Philip Hans Franses and Dick Van Dijk. Forecasting stock market volatility using (non-linear) Garch models. *Journal of Forecasting*, 15(3):229–235, 1996.

[GM85] Everette S. Gardner and Ed. Mckenzie. Forecasting Trends in Time Series. *Management Science*, 1985.

[Gne05] T. Gneiting. ATMOSPHERIC SCIENCE: Weather Forecasting with Ensemble Methods. *Science*, 310(5746):248–249, 2005.

[HNW05] Wei Huang, Yoshiteru Nakamori, and Shou Yang Wang. Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*, 32(10):2513–2522, 2005.

[Hol57] C. C Holt. Forecasting trends and seasonals by exponentially weighted moving averages. *ONR Memorandum*, 1957.

[Hyn] Rob J Hyndman. Forecasting : principles and practice.

[KKR06] Rakesh Kaundal, Amar A. Kapoor, and Gajendra P.S. Raghava. Machine learning techniques in disease forecasting: A case study on rice blast prediction. *BMC Bioinformatics*, 7:1–16, 2006.

[Kon08] Ai Kong. Retail Sales Forecast: A Cross Sectional Approach for Retail Investment Strategy. page 57, 2008.

[KY17] Yuta Kaneko and Katsutoshi Yada. A Deep Learning Approach for the Prediction of Retail Store Sales. *IEEE International Conference on Data Mining Workshops, ICDMW*, pages 531–537, 2017.

[LLLW11] Jeffrey K. Lazo, Megan Lawson, Peter H. Larsen, and Donald M. Waldman. U.S. Economic Sensitivity to Weather Variability. *Bulletin of the American Meteorological Society*, 92(6):709–720, 2011.

[MDMFPL10] Kyle B. Murray, Fabrizio Di Muro, Adam Finn, and Peter Popkowski Leszczyc. The effect of weather on consumer spending. *Journal of Retailing and Consumer Services*, 17(6):512–520, 2010.

[MFH16] Shaohui Ma, Robert Fildes, and Tao Huang. Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*, 249(1):245–257, 2016.

[MKF18] Shaohui Ma, Stephan Kolassa, and Robert Fildes. Retail forecasting: research and practice Evaluation and development of forecasting approaches applicable to Internet-based telecommunications View project forecasting in retail View project. (October), 2018.

[MP10] Marika Miklautsch and Marina Prem. *Sales promotions*, volume 12. 2010.

[MSA18] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3):1–26, 2018.

[Par10] Andrew G. Parsons. The Association Between Daily Weather and Daily Shopping Patterns. *Australasian Marketing Journal (AMJ)*, 2010.

[PE14] Dilek Penpece and Orhan Emre Elma. Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey. *International Journal of Trade, Economics and Finance*, 5(5):435–440, 2014.

[RSR15] Patrícia Ramos, Nicolau Santos, and Rui Rebelo. Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 2015.

[SB05] Aris A. Syntetos and John E. Boylan. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2):303–314, 2005.

[SD17] Robert Siwerz and Christopher Dahlen. Predicting sales in a food store department using machine learning. 2017.

[Sea18] Brian Seaman. Considerations of a retail forecasting practitioner. *International Journal of Forecasting*, 2018.

[SLP15] Suhartono, Muhammad Hisyam Lee, and Dedy Dwi Prastyo. Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects. *AIP Conference Proceedings*, 1691(December 2015), 2015.

[WCW+87] M. C. Weinstein, P. G. Coxson, L. W. Williams, T. M. Pass, W. B. Stason, and L. Goldman. Forecasting coronary heart disease incidence, mortality, and cost: The coronary heart disease policy model. *American Journal of Public Health*, 77(11):1417–1426, 1987.

[Win60] Peter R. Winters. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, 1960.

[WSH99] Rich Wolski, Neil T. Spring, and Jim Hayes. Network weather service: a distributed resource performance forecasting service for metacomputing. *Future Generation Computer Systems*, 15(5):757–768, 1999.

[XDH08] Xiangsheng Xie, Jiajun Ding, and Gang Hu. Forecasting the retail sales of china's catering industry using support vector machines. *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, (November 2005):4453–4457, 2008.

[ZQ05] G. Peter Zhang and Min Qi. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2):501–514, 2005.

Appendix A

Table total sales

section*

Table total sales

Ref	Abstract	Data	Conclusions
[AQS01]	A comparative study between neural networks and traditional forecast models.	US retail sales	ANN had a better performance, followed by ARIMA and Holt'Winters. Multivariate regression performed last.
[ABGM15]	Compared full seasonal models with dummy season. Also forecast multistep so the research is more close to a real environment.	South Africa retail sales	The models with dummy season variables had a better performance. Also any model distinguish from the other, some were better in booms and others in recession
[CZ03]	The purpose of this paper is to compare the accuracy of various linear and nonlinear models for forecasting aggregate retail sales	US retail sales	It modelled seasonality with trigonometric functions, deseason and dummy variables. Ann with deseason data had the best performance. Trigonometric functions are not usefull to model seasonality.
[ZQ05]	How to best model seasonality and trend using neural networks.	US retail sales;industrial production series; US total new privately owned housing units started from 1992 to 2001	It modelled seasonality with trigonometric functions, deseason and dummy variables. Ann with deseason data had the best performance. Trigonometric functions are not usefull to model seasonality.
[XDH08]	Compared support vector machines with Sarimax Model.	China Catering Sales	Support vector machines had a better performance then Sarimax.
Ref	Abstract	Data	Conclusions
[KY17]	The purpose of this research is to construct a sales prediction model for retail stores using the deep learning approach, which has gained significant attention in the rapidly developing field of machine learning in recent years.	3 years of pos data	The deep learning approach was compared to logistic regression, and the first a prediction accuracy 10 per cent greater..

Appendix B

Table Product Demand

Ref	Abstract	Variaveis	Conclusions
[SD17]	Which of the: MLP, RBFN and SVM achieves the highest accuracy when predicting the number of sold products in a food store department.	Day,year, month, holiday dummy, sum of all prices in the department	SVM had the best performance.
[ABGM15]	The objective of this study is to develop SARIMAX)model which tries to account all the effects due to the demand influencing factors, to forecast the daily sales of perishable foods in a retail store.	Seasonality, Holydays, Promotions Due to lack of data, cannibalization and weather effects are not taken into consideration.	Sarimax had better performance then Sarima.
[PE14]	forecast the sales revenue of grocery retailing industry in Turkey with the help of grocery retailers marketing costs, gross profit, and its competitors' gross profit by using artificial neural network	Marketing Costs, Gross Profit, Competitors Gross Profit	Not relevant

Table Product Demand

Ref	Abstract	Variaveis	Conclusions
[ASvWF09]	we propose using regression trees with explicit features constructed from sales and promotion time series of the focal and related SKU-store combination.	Statistical variables from data(average,sum,standart deviation) such as price, discount, advertise	On time without promotions simple time series has the best performance. On time with discount regression trees had better results.
[EÖK09]	The objective of the paper is to propose a new forecasting mechanism which is modeled by artificial intelligence approaches including the comparison of both arti-ficial neural networks and adaptive network-based fuzzy.	Product quality (1 to 9, evaluated by the costumers) Customer satisfaction,promotions, holidays and special days	Uses NN with neuro fuzzy model, and get a better result then both separated.
[BF09]	We use time series analysis to forecast the demand for beer on the Slovenian market using scanner data from two major retail stores Used Armax, autoregressive with exogenous variables.	Dummy New Year, Promotion , Temperature	Not relevant